

ATGGGTGCGAGAGCGTCAGTATTAAGCGGGGAGAATTAGATCGATGGGAAAAAAT  
TCGGTTAAGGCCAGGGGGAAAGAAGAAGTACAAGCTAAAGCACATCGTATGGGCAA  
GCAGGGAGCTAGAACGATTTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGC  
TGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAGGAGCT  
TCGATCACTATACAACACAGTAGCAACCCTCTATTGTGTGCACCAGCGGATCGAGA  
TCAAGGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAGTCCAAG  
AAGAAGGCCCAGCAGGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAA  
TTACCCTATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTA  
GAACTTTAAATGCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTG  
ATACCCATGTTTTTCAGCATTATCAGAAGGAGCCACCCACAGGACCTGAACACGAT  
GTTGAACACCGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGACCATCA  
ATGAGGAAGCTGCAGAATGGGATAGAGTGCATCCAGTGCATGCAGGGCCTATTGCA  
CCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGAACTACTAGTACCCT  
TCAGGAACAAATAGGATGGATGACAAATAATCCACCTATCCCAGTAGGAGAGATCT  
ACAAGAGGTGGATAATCCTGGGATTGAACAAGATCGTGAGGATGTATAGCCCTACC  
AGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGACTATGTAGACCG  
GTTCTATAAACTCTAAGAGCTGAGCAAGCTTCACAGGAGGTAAAAAATTGGATGA  
CAGAAACCTTGTTGGTCCAAAATGCGAACCAGATTGTAAGACCATCCTGAAGGCT  
CTCGGCCCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGG  
ACCCGGCCATAAGGCAAGAGTTTTGGCCGAGGCGATGAGCCAGGTGACGAACTCGG  
CGACCATAATGATGCAGAGAGGCAACTTCCGGAACCAGCGGAAGATCGTCAAGTGC  
TTCAATTGTGGCAAAGAAGGGCACACCGCCAGGAACTGCCGGGCCCCCGGAAGAA  
GGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGAC

FIG. 1

AGGCTAATTTTTTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTT  
CTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTCTGGGGT  
AGAGACAACAACCTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAACCTGTATCCTT  
TAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGGATCGGGG  
GGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAGAA  
ATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGGATCGGGGGCTTCAT  
CAAGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGTGGACATAAAGCTATAG  
GTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACC  
CAGATCGGCTGCACCTTGAACCTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAA  
GTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGA  
AGATCAAGGCCTTAGTCGAAATCTGTACAGAGATGGAGAAGGAAGGGAAGATCAGC  
AAGATCGGGCCTGAGAACCCCTACAACACTCCAGTCTTCGCAATCAAGAAGAAGGA  
CAGTACCAAGTGGAGAAAGCTGGTGGACTTCAGAGAGCTGAACAAGAGAACTCAGG  
ACTTCTGGGAAGTTCAGCTGGGCATCCCACATCCCGCTGGGTGAAGAAGAAGAAG  
TCAGTGACAGTGCTGGATGTGGGTGATGCCTACTTCTCCGTTCCCTTGGACGAGGA  
CTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCAACAACGAGACACCAGGCA  
TCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGATCACCAGCCATCTTT  
CAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAACCCAGACATCGT  
GATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCGGGCAGC  
ACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCACA  
CCAGACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACTGCA  
TCCTGACAAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTG  
TGAACGACATACAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCA  
GGCATCAAAGTTAGGCAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGA

AGTGATCCCACTGACAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCC  
TGAAGGAGCCAGTACATGGAGTGACTACGACCCAAGCAAGGACCTGATCGCAGAG  
ATCCAGAAGCAGGGGCAAGGCCAATGGACCTACCAAATCTACCAGGAGCCCTTCAA  
GAACCTGAAGACAGGCAAGTACGCAAGGATGAGGGGTGCCACACCAACGATGTGA  
AGCAGCTGACAGAGGCAGTGCAGAAGATCACCACAGAGAGCATCGTGATCTGGGGC  
AAGACTCCCAAGTTCAAGCTGCCCATACAGAAGGAGACATGGGAGACATGGTGGAC  
CGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTCGTGAACACCCCTCCCT  
TGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGGGAGCAGAGACCTTC  
TACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAAGGCAGGCTACGTGAC  
CAACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAACCAGAAGACTG  
AGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCGTG  
ACAGACTCACAGTACGCACTGGGCATCATCCAAGCACAACCAGACCAATCCGAGTC  
AGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTAACCTGG  
CATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTC  
AGTGCTGGGATCCGGAAGGTGCTGTTCCCTGGACGGGATCGATAAGGCCCAAGATGA  
ACATGAGAAGTACCACTCCAACCTGGCGCGCTATGGCCAGCGACTTCAACCTGCCAC  
CTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAA  
GCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAGCTGGACTGCACGCA  
CCTGGAGGGGAAGGTGATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAG  
CAGAAGTTATCCCTGCTGAAACTGGGCAGGAAACAGCATATTTTCTTTTAAATTA  
GCAGGAAGATGGCCAGTAAAAACAATACACACGGACAACGGAAGCAACTTCACTGG  
TGCTACGGTTAAGGCCGCCTGTTGGTGGGCGGGAATCAAGCAGGAATTTGGAATTC  
CCTACAATCCCCAATCGCAAGGAGTCGTGGAGAGCATGAACAAGGAGCTGAAGAAG  
ATCATCGGACAAGTGAGGGATCAGGCTGAGCACCTGAAGACAGCAGTGCAGATGGC

AGTGTTTCATCCACAACCTTCAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGG  
AAAGGATCGTGGACATCATCGCCACCGACATCCAAACCAAGGAGCTGCAGAAGCAG  
ATCACCAAGATCCAGAACTTCCGGGTGTACTACCGCGACAGCCGCAACCCACTGTG  
GAAGGGACCAGCAAAGCTCCTCTGGAAGGGAGAGGGGGCAGTGGTGATCCAGGACA  
ACAGTGACATCAAAGTGGTGCCAAGGCGCAAGGCCAAGATCATCCGCGACTATGGA  
AAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATTAGAACCT  
GGAAGAGCCTGGTGAAGCACCATATG (SEQUENCE ID NO:1)

>wildtype	TGTACAGAGA TGGAAAAGGA AGGGAAAATT TCAAAAATTG
>mutated	TGTACAGAGA TGGAGAAGGA AGGGAAAGATC AGCAAGATCG
#1	.....*
>wildtype	GGCCTGAAAA TCCATACAAT ACTCCAGTAT TTGCCATAAA
>mutated	GGCCTGAGAA CCCCTACAAC ACTCCAGTCT TCGCAATCAA
#41	.....*
>wildtype	GAAAAAAGAC AGTACTAAAT GGAGAAAATT AGTAGATTTC
>mutated	GAAGAAGGAC AGTACCAAGT GGAGAAAGCT GGTGGACTTC
#81	.....*
>wildtype	AGAGAACTTA ATAAGAGAAC TCAAGACTTC TGGGAAGTTC
>mutated	AGAGAGCTGA ACAAGAGAAC TCAGGACTTC TGGGAAGTTC
#121	.....*
>wildtype	AATTAGGAAT ACCACATCCC GCAGGGTTAA AAAAGAAAAA
>mutated	AGCTGGGCAT CCCACATCCC GCTGGGTTGA AGAAGAAGAA
#161	.....*
>wildtype	ATCAGTAACA GTACTGGATG TGGGTGATGC ATATTTTTC
>mutated	GTCAGTGACA GTGCTGGATG TGGGTGATGC CTACTTCTCC
#201	.....*
>wildtype	GTTCCCTTAG ATGAAGACTT CAGGAAATAT ACTGCATTTA
>mutated	GTTCCCTTGG ACGAGGACTT CAGGAAGTAC ACTGCCTTCA
#241	.....*
>wildtype	CCATACCTAG TATAACAAT GAGACACCAG GGATTAGATA
>mutated	CGATACCTAG CATCAACAAC GAGACACCAG GCATCCGCTA
#281	.....*
>wildtype	TCAGTACAAT GTGCTTCCAC AGGGATGGAA AGGATCACCA
>mutated	CCAGTACAAC GTGCTGCCAC AGGGATGGAA GGGATCACCA
#321	.....*
>wildtype	GCAATATTCC AAAGTAGCAT GACAAAAATC TTAGAGCCTT
>mutated	GCCATCTTTC AAAGCAGCAT GACCAAGATC CTGGAGCCCT
#361	.....*
>wildtype	TTAGAAAACA AAATCCAGAC ATAGTTATCT ATCAATACAT
>mutated	TCCGCAAGCA AAACCCAGAC ATCGTGATCT ATCAGTACAT
#401	.....*

FIG. 2

```

>wildtype      GGATGATTTG TATGTAGGAT CTGACTTAGA AATAGGGCAG
>mutated       GGACGACCTC TACGTAGGAA GTGACCTGGA GATCGGGCAG
#441           .....
               *  *  *  *  *  *  *  *

>wildtype      CATAGAACAA AAATAGAGGA GCTGAGACAA CATCTGTTGA
>mutated       CACAGGACCA AGATCGAGGA GCTGAGACAG CATCTGTTGA
#481           .....
               *  *  *  *  *

>wildtype      GGTGGGGACT TACCACACCA GACAAAAAAC ATCAGAAAGA
>mutated       GGTGGGGACT GACCACACCA GACAAGAAGC ACCAGAAGGA
#521           .....
               *  *  *  *

>wildtype      ACCTCCATTG CTTTGGATGG GTTATGAACT CCATCCTGAT
>mutated       ACCTCCCTTC CTGTGGATGG GCTACGAACT GCATCCTGAC
#561           .....
               *  *  *  *

>wildtype      AAATGGACAG TACAGCCTAT AGTGCTGCCA GAAAAAGACA
>mutated       AAGTGGACAG TGCAGCCCAT CGTGCTGCCT GAGAAGGACA
#601           .....
               *  *  *  *

>wildtype      GCTGGACTGT CAATGACATA CAGAAGTTAG TGGGGAAATT
>mutated       GCTGGACTGT GAACGACATA CAGAAGCTCG TGGGCAAGTT
#641           .....
               *  *  *  *

>wildtype      GAATTGGGCA AGTCAGATT ACCCAGGGAT TAAAGTAAGG
>mutated       GAAGTGGGCA AGCCAGATCT ACCCAGGCAT CAAAGTTAGG
#681           .....
               *  *  *  *

>wildtype      CAATTATGTA AACTCCTTAG AGGAACCAAA GCACTAACAG
>mutated       CAGCTGTGCA AGCTGCTTCG AGGAACCAAG GCACTGACAG
#721           .....
               *  *  *  *

>wildtype      AAGTAATACC ACTAACAGAA GAAGCAGAGC TAGAACTGGC
>mutated       AAGTGATCCC ACTGACAGAG GAAGCAGAGC TAGAACTGGC
#761           .....
               *  *  *  *

>wildtype      AGAAAACAGA GAGATTCTAA AAGAACCAGT ACATGGAGTG
>mutated       AGAGAACCGA GAGATCCTGA AGGAGCCAGT ACATGGAGTG
#801           .....
               *  *  *  *

>wildtype      TATTATGACC CATCAAAGA CTTAATAGCA GAAATACAGA
>mutated       TACTACGACC CAAGCAAGGA CCTGATCGCA GAGATCCAGA
#841           .....
               *  *  *  *

```

```

>wildtype      AGCAGGGGCA AGGCCAATGG ACATATCAAA TTTATCAAGA
>mutated       AGCAGGGGCA AGGCCAATGG ACCTACCAA TCTACCAGGA
#881           .....
                        * * * * *

>wildtype      GCCATTTAAA AATCTGAAAA CAGGAAAATA TGCAAGAATG
>mutated       GCCCTTCAAG AACCTGAAGA CAGGCAAGTA CGCAAGGATG
#921           .....
                        * * * * *

>wildtype      AGGGGTGCCC ACACTAATGA TGTAACAA TTAACAGAGG
>mutated       AGGGGTGCCC ACACCAACGA TGTGAAGCAG CTGACAGAGG
#961           .....
                        * * * * *

>wildtype      CAGTGCAAAA AATAACCACA GAAAGCATAG TAATATGGGG
>mutated       CAGTGCAGAA GATCACCACA GAGAGCATCG TGATCTGGGG
#1001          .....
                        * * * * *

>wildtype      AAAGACTCCT AAATTTAAAC TGCCCATACA AAAGGAAACA
>mutated       CAAGACTCCC AAGTTCAAGC TGCCCATACA GAAGGAGACA
#1041          .....
                        * * * * *

>wildtype      TGGGAAACAT GGTGGACAGA GTATTGGCAA GCCACCTGGA
>mutated       TGGGAGACAT GGTGGACCGA GTACTGGCAA GCCACCTGGA
#1081          .....
                        * * *

>wildtype      TTCCTGAGTG GGAGTTTGTT AATACCCCTC CTTTAGTGAA
>mutated       TCCCTGAGTG GGAGTTCGTG AACACCCCTC CCTTGGTGAA
#1121          .....
                        * * * * *

>wildtype      ATTATGGTAC CAGTTAGAGA AAGAACCCAT AGTAGGAGCA
>mutated       ACTGTGGTAT CAGCTGGAGA AGGAACCCAT CGTGGGAGCA
#1161          .....
                        * * * * *

>wildtype      GAAACCTTCT ATGTAGATGG GGCAGCTAAC AGGGAGACTA
>mutated       GAGACCTTCT ACGTGGATGG GGCAGCCAAC AGGGAGACCA
#1201          .....
                        * * * * *

>wildtype      AATTAGGAAA AGCAGGATAT GTTACTAATA GAGGAAGACA
>mutated       AGCTGGGCAA GGCAGGCTAC GTGACCAACC GAGGACGACA
#1241          .....
                        * * * * *

>wildtype      AAAAGTTGTC ACCCTAACTG ACACAACAAA TCAGAAGACT
>mutated       GAAAGTGGTG ACCCTGACTG ACACCACCAA CCAGAAGACT
#1281          .....
                        * * * * *

```

```

>wildtype      GAGTTACAAG CAATTTATCT AGCTTTGCAG GATTCGGGAT
>mutated       GAGCTGCAAG CCATCTACCT AGCTCTGCAA GACAGCGGAC
#1321          .....
               * * * * *

>wildtype      TAGAAGTAAA CATAGTAACA GACTCACAAT ATGCATTAGG
>mutated       TGGAAGTGAA CATCGTGACA GACTCACAGT ACGCACTGGG
#1361          .....
               * * * * *

>wildtype      AATCATTCAA GCACAACCAG ATCAAAGTGA ATCAGAGTTA
>mutated       CATCATCCAA GCACAACCAG ACCAATCCGA GTCAGAGCTG
#1401          .....
               * * * * *

>wildtype      GTCAATCAAA TAATAGAGCA GTTAATAAAA AAGGAAAAGG
>mutated       GTGAACCAGA TCATCGAGCA GCTGATCAAG AAGGAGAAAAG
#1441          .....
               * * * * *

>wildtype      TCTATCTGGC ATGGGTACCA GCACACAAAG GAATTGGAGG
>mutated       TGTACCTGGC ATGGGTACCA GCACACAAAG GAATTGGAGG
#1481          .....
               * *

>wildtype      AAATGAACAA GTAGATAAAT TAGTCAGTGC TGGGAATCAGG
>mutated       AAATGAACAA GTAGATAAAT TAGTCAGTGC TGGGATCCGG
#1521          .....
               * *

>wildtype      AAAGTACTAT TTTTAGATGG AATAGATAAG GCCCAAGATG
>mutated       AAGGTGCTGT TCCTGGACGG GATCGATAAG GCCCAAGATG
#1561          .....
               * * * * *

>wildtype      AACATGAGAA ATATCACAGT AATTGGAGAG CAATGGCTAG
>mutated       AACATGAGAA GTACCACTCC AACTGGCGCG CTATGGCCAG
#1601          .....
               * * * * *

>wildtype      TGATTTTAAC CTGCCACCTG TAGTAGCAAA AGAAATAGTA
>mutated       CGACTTCAAC CTGCCACCTG TAGTAGCAAA AGAAATAGTA
#1641          .....
               * * *

>wildtype      GCCAGCTGTG ATAAATGTCA GCTAAAAGGA GAAGCCATGC
>mutated       GCCAGCTGTG ATAAATGTCA GCTAAAAGGA GAAGCCATGC
#1681          .....

>wildtype      ATGGACAAGT AGACTGTAGT CCAGGAATAT GGCAACTAGA
>mutated       ATGGACAAGT AGACTGTAGT CCAGGAATAT GGCAGCTGGA
#1721          .....
               * *

```

```

>wildtype      TTGTACACAT TTAGAAGGAA AAGTTATCCT GGTAGCAGTT
>mutated       CTGCACGCAC CTGGAGGGGA AGGTGATCCT GGTAGCAGTT
#1761          .....
               * * * * *

>wildtype      CATGTAGCCA GTGGATATAT AGAAGCAGAA GTTATTCCAG
>mutated       CATGTAGCCA GTGGATATAT AGAAGCAGAA GTTATCCCTG
#1801          .....
                                   * *

>wildtype      CAGAAACAGG GCAGGAAACA GCATATTTTC TTTTAAAATT
>mutated       CTGAAACTGG GCAGGAAACA GCATATTTTC TTTTAAAATT
#1841          .....
               * *

>wildtype      AGCAGGAAGA TGGCCAGTAA AAACAATACA TACAGACAAT
>mutated       AGCAGGAAGA TGGCCAGTAA AAACAATACA CACGGACAAC
#1881          .....
                                   * * *

>wildtype      GGCAGCAATT TCACCAGTGC TACGGTTAAG GCCGCCTGTT
>mutated       GGAAGCAACT TCACTGGTGC TACGGTTAAG GCCGCCTGTT
#1921          .....
               * * *

>wildtype      GGTGGGCGGG AATCAAGCAG GAATTTGGAA TTCCCTACAA
>mutated       GGTGGGCGGG AATCAAGCAG GAATTTGGAA TTCCCTACAA
#1961          .....

>wildtype      TCCCCAAAGT CAAGGAGTAG TAGAATCTAT GAATAAAGAA
>mutated       TCCCCAATCG CAAGGAGTCG TGGAGAGCAT GAACAAGGAG
#2001          .....
               *** * * * *

>wildtype      TTAAAGAAAA TTATAGGACA GGTAAGAGAT CAGGCTGAAC
>mutated       CTGAAGAAGA TCATCGGACA AGTGAGGGAT CAGGCTGAGC
#2041          .....
               * * * * *

>wildtype      ATCTTAAGAC AGCAGTACAA ATGGCAGTAT TCATCCACAA
>mutated       ACCTGAAGAC AGCAGTGCAG ATGGCAGTGT TCATCCACAA
#2081          .....
               * * * *

>wildtype      TTTTAAAAGA AAAGGGGGGA TTGGGGGGTA CAGTGCAGGG
>mutated       CTTCAAAAAGA AAAGGGGGGA TTGGGGGGTA CAGTGCAGGG
#2121          .....
               * *

>wildtype      GAAAGAATAG TAGACATAAT AGCAACAGAC ATACAAACTA
>mutated       GAAAGGATCG TGGACATCAT CGCCACCGAC ATCCAAACCA
#2161          .....
               * * * * *

```

```

>wildtype      AAGAATTACA AAAACAAATT ACAAAAATTC AAAATTTTCG
>mutated       AGGAGCTGCA GAAGCAGATC ACCAAGATCC AGAACTTCCG
#2201          .....
               *  *  *  *  *  *  *  *  *  *  *  *

>wildtype      GGTTTATTAC AGGGACAGCA GAAATCCACT TTGGAAAGGA
>mutated       GGTGTACTAC CGCGACAGCC GCAACCCACT GTGGAAGGGA
#2241          .....
               *  *  *  *  *  *  *  *  *  *

>wildtype      CCAGCAAAGC TCCTCTGGAA AGGTGAAGGG GCAGTAGTAA
>mutated       CCAGCAAAGC TCCTCTGGAA GGGAGAGGGG GCAGTGGTGA
#2281          .....
               *  *  *  *  *  *  *  *  *  *

>wildtype      TACAAGATAA TAGTGACATA AAAGTAGTGC CAAGAAGAAA
>mutated       TCCAGGACAA CAGTGACATC AAAGTGGTGC CAAGGCGCAA
#2321          .....
               *  *  *  *  *  *  *  *  *  *

>wildtype      AGCAAAGATC ATTAGGGATT ATGGAAAACA GATGGCAGGT
>mutated       GGCCAAGATC ATCCGCGACT ATGGAAAACA GATGGCAGGT
#2361          .....
               *  *  *  *  *  *  *  *  *  *

>wildtype      GATGATTGTG TGGCAAGTAG ACAGGATGAG GATTAGAACA
>mutated       GATGATTGTG TGGCAAGTAG ACAGGATGAG GATTAGAACC
#2401          .....
               *  *  *  *  *  *  *  *  *  *

>wildtype      TGGAAAAGTT TAGTAAAACA CCATATG
>mutated       TGGAAGAGCC TGGTGAAGCA CCATATG
#2441          .....
               *  *  *  *  *  *  *  *  *  *

```

ATGGGCGTGAGAACTCCGTCTTGT CAGGGAAGAAAGCAGATGAATTAG  
AAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGC  
ATGTAGTATGGGCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAG  
CCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTTCGGTCTTAGCT  
CCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG  
TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGA  
GGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAAC  
CACCGAAACCATGCCGAAGACCTCTCGACCAACAGCACCATCTAGCGGC  
AGAGGAGGAAACTACCCAGTACAGCAGATCGGTGGCAACTACGTCCAC  
CTGCCACTGTCCCCGAGAACCCTGAACGCTTGGGTCAAGCTGATCGAGG  
AGAAGAAGTTCGGAGCAGAAGTAGTGCCAGGATTCCAGGCACTGTCAG  
AAGGTTGCACCCCCTACGACATCAACCAGATGCTGAACTGCGTTGGAGA  
CCATCAGGCGGCTATGCAGATCATCCGTGACATCATCAACGAGGAGGCT  
GCAGATTGGGACTTGCAGCACCCACAACCAGCTCCACAACAAGGACAA  
CTTAGGGAGCCGTCAGGATCAGACATCGCAGGAACCACTCCTCAGTTG  
ACGAACAGATCCAGTGGATGTACCGTCAGCAGAACCCGATCCCAGTAGG  
CAACATCTACCGTCGATGGATCCAGCTGGGTCTGCAGAAATGCGTCCGT  
ATGTACAACCCGACCAACATTCTAGATGTAAAACAAGGGCCAAAAGAG  
CCATTT CAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAAC  
AGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACTGCTGATTCA  
AAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAAT  
CCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGGCCG  
GGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCA  
CCAGTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGC  
CAATTAAGTGTTGGAATTGTGGGAAAGAGGGGACACTCTGCAAGGCAATG  
CAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACCA  
TGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGT  
CCATGGGGAAAGAAGCCCCGCAATTTCCCATGGCTCAAGTGCATCAGG  
GGCTGATGCCAACTGCTCCCCAGAGGACCCAGCTGTGGATCTGCTAAA  
GAACTACATGCAGTTGGGCAAGCAGCAGAGAGAAAGCAGAGAGAAAG  
CAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACCTCAAT  
TCTCTCTTTGGAGGAGACCAAGTAG

FIG. 3

```

SIV gag  -----
#1      .....
      ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAG

SIV gag  -----
#41     .....
      ATGAATTAGAAAAAATTAGGCTACGACCCAACGGAAAGAA

SIV gag  -----
#81     .....
      AAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA

SIV gag  -----
#121    .....
      TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACA

SIV gag  -----
#161    .....
      AAGAAGGATGTCAAAAAATACTTTCGGTCTTAGCTCCATT

SIV gag  -----
#201    .....
      AGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT

SIV gag  -----
#241    .....
      ACTGTCTGCGTCATCTGGTGCATTACGCAGAAGAGAAAG

SIV gag  -----
SIVgagDX.. -----
#281    .....
      TGAAACACACTGAGGAAGCAAAACAGATAGTGCAGAGACA

SIV gag  -----A--A----T----A--A
SIVgagDX..-----C--C----C----G--G
#321    .....
      CCTAGTGGTGGAAACAGGAACMACMGAAACYATGCCRAAR

SIV gag  --AAG-A-----
SIVgagDX..--CTC-C-----
#361    .....
      ACMWSTMGACCAACAGCACCATCTAGCGGCAGAGGAGGAA

```

FIG. 4

SIV gag    -T-----A--A--A-----T-----  
 SIVgagDX..-C-----G--G--C-----C-----  
 #401       .....  
           AYTACCCAGTACARCARATMGGTGGTAACTAYGTCCACCT

SIV gag    ----T-AAG-----AT-A--T--C-----A--AT--  
 SIVgagDX..----C-GTC-----CC-G--C--T-----C--GC--  
 #441       .....  
           GCCAYTRWSCCCGAGAACMYTRAAYGCTGGGTMAARYTG

SIV gag    --A-----A-----A--T-----  
 SIVgagDX..--C-----G-----G--C-----  
 #481       .....  
           ATMGAGGARAAGAARTTYGGAGCAGAAGTAGTGCCAGGAT

SIV gag    -T-----T-----T-----  
 SIVgagDX..-C-----C-----C-----  
 #521       .....  
           TYCAGGCACTGTCAGAAGGTTGCACCCCTAYGACATYAA

SIV gag    T-----T-A--T--T--G-----A-----  
 SIVgagDX..C-----C-G--C--C--T-----G-----  
 #561       .....  
           YCAGATGYTRAAYTGYGTKGGAGACCATCARGCGGCTATG

SIV gag    ----T---A-A--T--T--A-----  
 SIVgagDX..----C---C-T--C--C--C-----  
 #601       .....  
           CAGATYATCMGWGAYATYATMAACGAGGAGGCTGCAGATT

SIV gag    -----  
 SIVgagDX..-----  
 #641       .....  
           GGGACTTGCAGCACCCACAACCAGCTCCACAACAAGGACA

SIV gag    -----T--T-----A--T  
 SIVgagDX..-----C--C-----C--C  
 #681       .....  
           ACTTAGGGAGCCGTCAGGATCAGAYATYGCAGGAACMACY

SIV gag    AGT-----A--T-----A-----A--A--A--  
 SIVgagDX..TCC-----T--C-----G-----C--T--G--  
 #721       .....  
           WSYTCAGTWGAYGAACARATCCAGTGGATGTACMGWCARC

SIV gag -----C--A-----T---A-GA-----  
 SIVgagDX..-----G--C-----C---C-TC-----  
 #761 .....  
 AGAACCCSATMCCAGTAGGCAACATYTACMGKMGATGGAT

SIV gag ---A-----GT----A--A--T--CA-A-----T-----A  
 SIVgagDX..---G-----TC----G--G--C--TC-T-----C-----G  
 #801 .....  
 CCARCTGGGKYTGCARAARTGYGYMGWATGTAYAACCCR

SIV gag --A-----  
 SIVgagDX..--C-----  
 #841 .....  
 ACMAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCAT

SIV gag -----  
 #881 .....  
 TTCAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGC

SIV gag -----  
 #921 .....  
 AGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAA

SIV gag -----  
 #961 .....  
 ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAG

SIV gag -----  
 #1001 .....  
 TGCTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAAT

SIV gag -----  
 #1041 .....  
 GCTGACGGCTTGTC AAGGAGTAGGGGGGCGGGACAGAAG

SIV gag -----  
 #1081 .....  
 GCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCAC

SIV gag -----  
 #1121 .....  
 CAGTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACC

SIV gag -----  
 #1161 .....  
 AAGAAAGCCAATTAAGTGTGGAATTGTGGGAAGAGGGGA

SIV gag -----  
#1201 .....  
CACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGAT

SIV gag -----  
#1241 .....  
GCTGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATG

SIV gag -----  
#1281 .....  
CCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGG

SIV gag -----  
#1321 .....  
GGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATC

SIV gag -----  
#1361 .....  
AGGGGCTGATGCCAACTGCTCCCCAGAGGACCCAGCTGT

SIV gag -----  
#1401 .....  
GGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG

SIV gag -----  
#1441 .....  
AGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGG

SIV gag -----  
#1481 .....  
AGGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGG

SIV gag -----  
#1521 .....  
AGGAGACCAGTAG

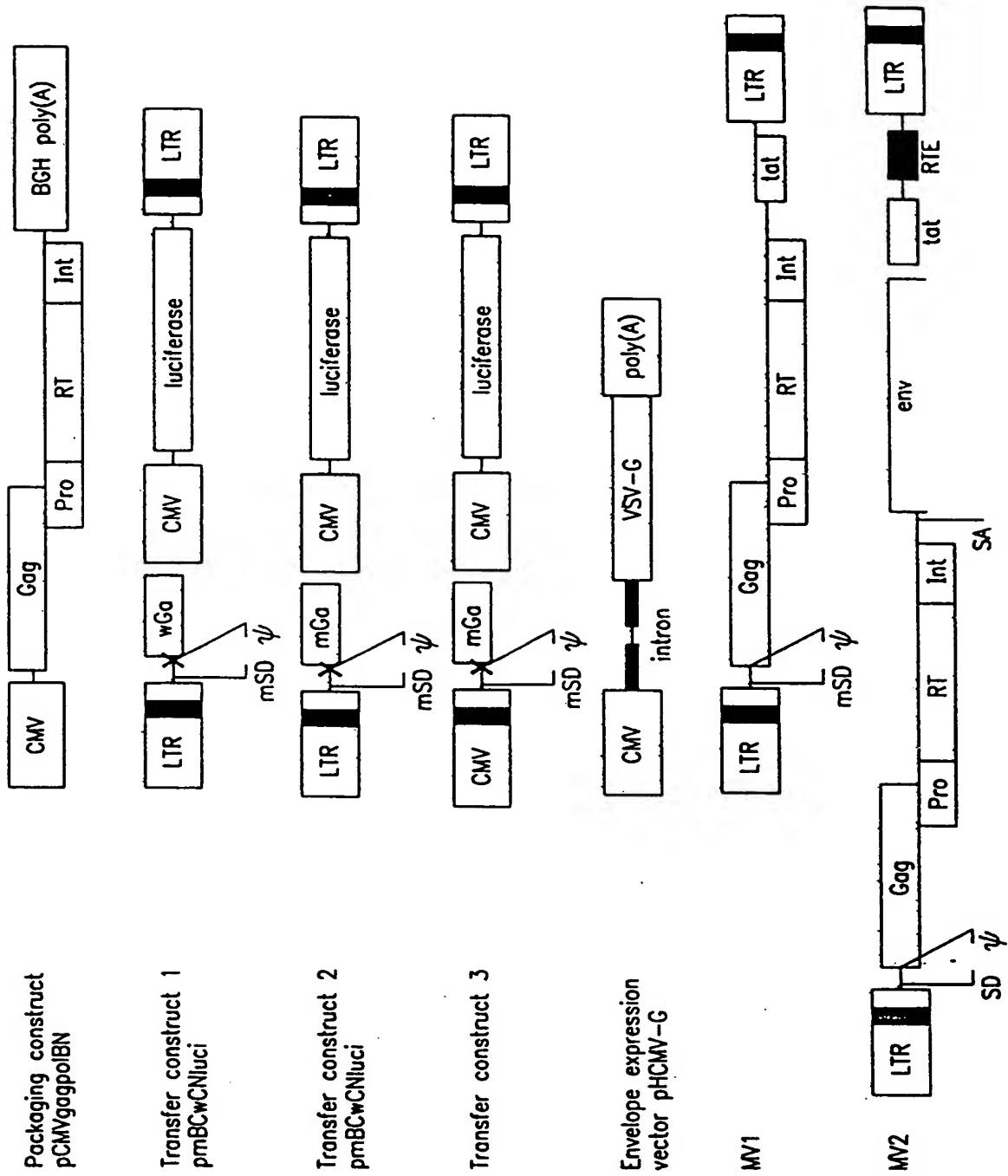


FIG. 5

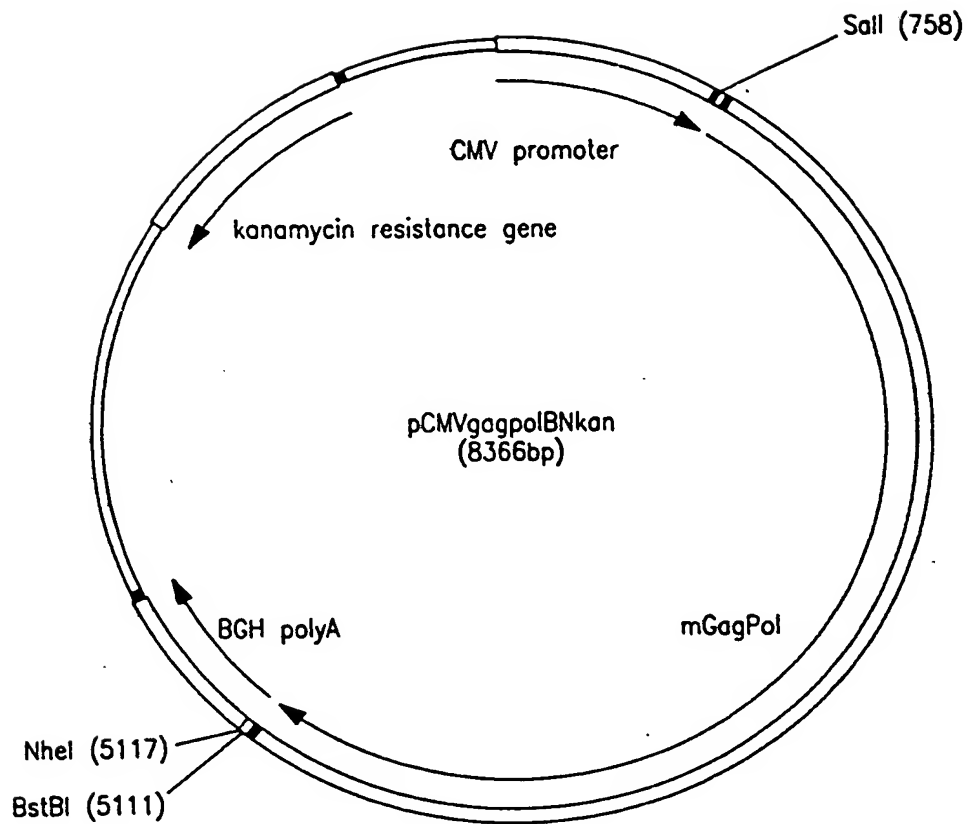


FIG. 6

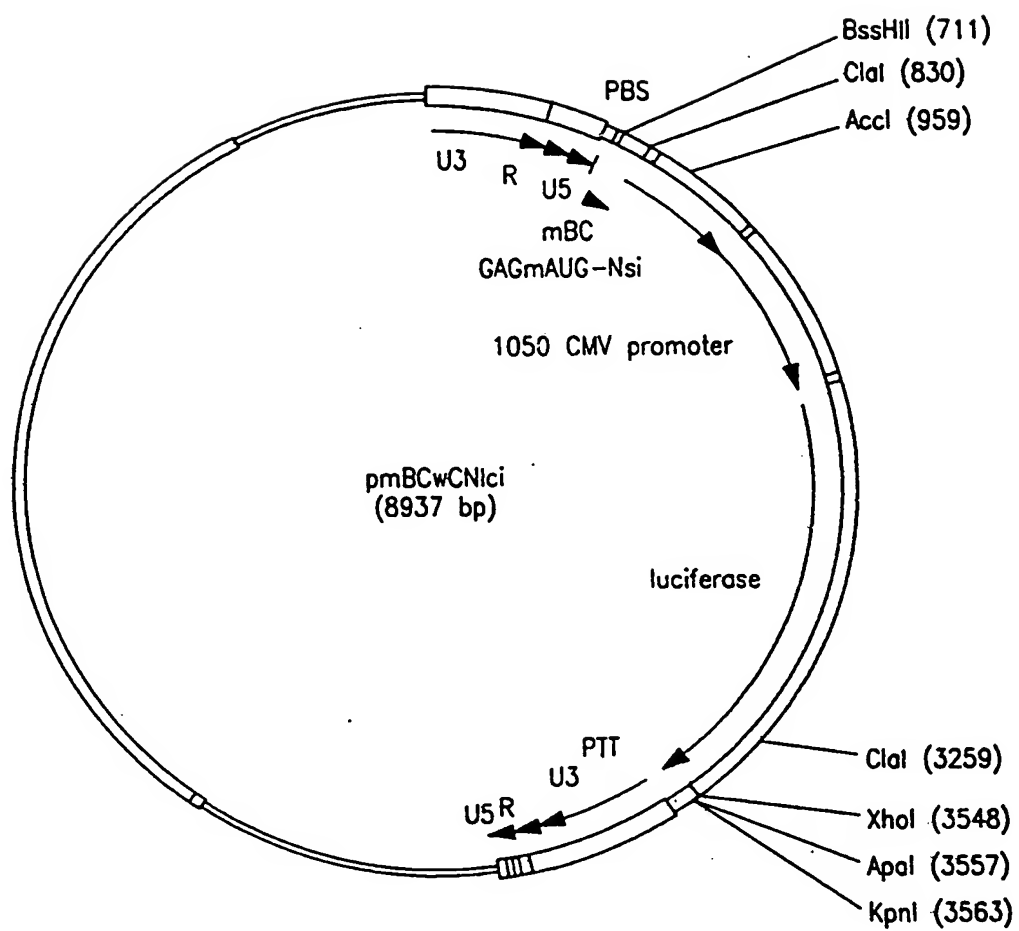


FIG. 7

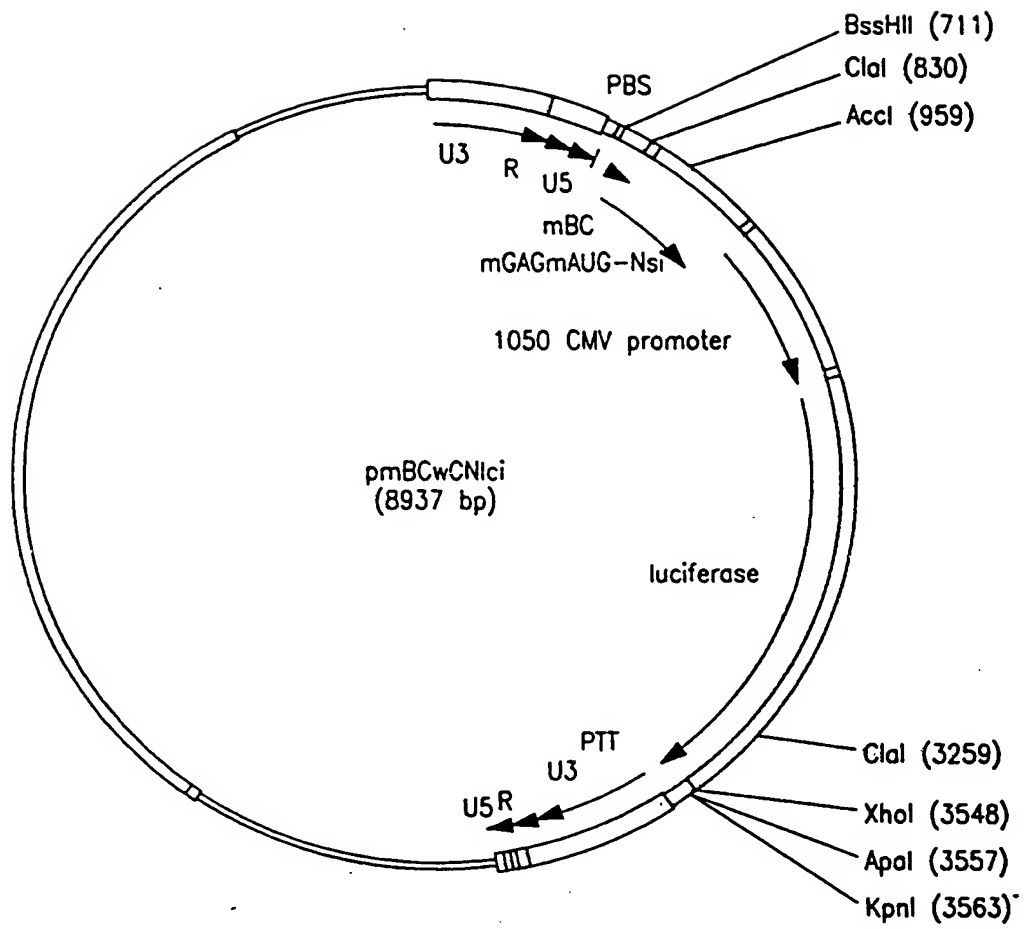


FIG. 8

```

1  CCTGGCCATT GCATACGTTG TATCCATATC ATAATATGTA CATTTATATT GGCTCATGTC CAACATTACC
71  GCCATGTTGA CATTGATTAT TGA CTAGTTA TTAATAGTAA TCAATTACGG GGTCAATTAGT TCATAGCCCA
141 TATATGGAGT TCCGCGTTAC ATAACTTACG GTAAATGGCC CGCCTGGCTG ACCGCCCAAC GACCCCGGCC
211 CATTGACGTC AATAATGACG TATGTTCCCA TAGTAACGCC AATAGGGACT TTCCATTGAC GTCAATGGGT
281 GGAGTATTTA CGGTAAACTG CCCACTTGGC AGTACATCAA GTGTATCATA TGCCAAGTAC GCCCCCTATT
351 GACGTCAATG ACGGTAAATG GCGCGCCTGG CATTATGCCC AGTACATGAC CTTATGGGAC TTTCTACTT
421 GGCAGTACAT CTACGTATTA GTCATCGCTA TTACCATGGT GATGCGGTTT TGGCAGTACA TCAATGGGCG
491 TGGATAGCGG TTGACTCAC GGGGATTTC AAGTCTCCAC CCCATTGACG TCAATGGGAG TTTGTTTTGG
561 CACCAAAATC AACGGGACTT TCCAAAATGT CGTAACAACT CCGCCCCATT GACGCAAATG GGCGGTAGGC
631 GTGTACGGTG GGAGGTCTAT ATAAGCAGAG CTCGTTTAGT GAACCGTCAG ATCGCCTGGA GACGCCATCC

701 ACGCTGTTTT GACCTCCATA GAAGACACCG GGACCGATCC AGCCTCCGGG SalI GCGCGCGGTC GACAGAGAGA (758)
771 TGGGTGCGAG AGCGTCAGTA TTAAGCGGGG GAGAATTAGA TCGATGGGAA AAAATTTCGGT TAAGGCCAGG
841 GGGAAAGAAG AAGTACAAGC TAAAGCACAT CGTATGGGCA AGCAGGGAGC TAGAACGATT CGCAGTTAAT
911 CCTGGCCTGT TAGAAACATC AGAAGGCTGT AGACAAATAC TGGGACAGCT ACAACCATCC CTTCAGACAG
981 GATCAGAGGA GCTTCGATCA CTATACAACA CAGTAGCAAC CCTCTATTGT GTGCACCAGC GGATCGAGAT
1051 CAAGGACACC AAGGAAGCTT TAGACAAGAT AGAGGAAGAG CAAAACAAGT CCAAGAAGAA GGCCAGCAG
1121 GCAGCAGCTG ACACAGGACA CAGCAATCAG GTCAGCCAAA ATTACCCTAT AGTGCAGAAC ATCCAGGGGC
1191 AAATGGTACA TCAGGCCATA TCACCTAGAA CTTTAAATGC ATGGGTAAAA GTAGTAGAAG AGAAGGCTTT
1261 CAGCCCAGAA GTGATACCCA TGTTTTCAGC ATTATCAGAA GGAGCCACCC CACAGGACCT GAACACGATG
1331 TTGAACACCG TGGGGGGACA TCAAGCAGCC ATGCAAATGT TAAAAGAGAC CATCAATGAG GAAGCTGCAG
1401 AATGGGATAG AGTGCATCCA GTGCATGCAG GGCCTATTGC ACCAGGCCAG ATGAGAGAAC CAAGGGGAAG
1471 TGACATAGCA GGAATACTA GTACCCTTCA GGAACAAATA GGATGGATGA CAAATAATCC ACCTATCCCA
1541 GTAGCAGAGA TCTACAAGAG GTGGATAATC CTGGGATTGA ACAAGATCGT GAGGATGTAT AGCCCTACCA
1611 GCATTCTGGA CATAAGACAA GGACCAAAGG AACCTTTAG AGACTATGTA GACCGGTTCT ATAAACTCT
1681 AAGAGCTGAG CAAGCTTCAC AGGAGGTAAA AAATTGGATG ACAGAAACCT TGTGGTCCA AAATGCGAAC
1751 CCAGATTGTA AGACCATCCT GAAGGCTCTC GGCCAGCGG CTACACTAGA AGAAATGATG ACAGCATGTC
1821 AGGGAGTAGG AGGACCCGGC CATAAGGCAA GAGTTTTGGC CGAGGCGATG AGCCAGGTGA CGAACTCGGC

```

FIG. 9A

1891 GACCATAATG ATGCAGAGAG GCAACTTCCG GAACCAGCGG AAGATCGTCA AGTGCTTCAA TTGTGCCAAA  
 1961 GAAGGGCACA CCGCCAGGAA CTGCCGGGCC CCCCAGGAAGA AGGGCTGTTG GAAATGTGGA AAGGAAGGAC  
 2031 ACCAAATGAA AGATTGTACT GAGAGACAGG CTAATTTTTT AGGGAAGATC TGGCCTTCCT ACAAGGGAAG  
 2101 GCCAGGGAAT TTTCTTCAGA GCAGACCAGA GCCAACAGCC CCACCAGAAG AGAGCTTCAG GTCTGGGGTA  
 2171 GAGACAACAA CTCCCCCTCA GAAGCAGGAG CCGATAGACA AGGAACTGTA TCCTTTAACT TCCCTCAGAT  
 2241 CACTCTTTGG CAACGACCCC TCGTCACAGT AAGGATCGGG GGGCAACTCA AGGAAGCGCT GCTCGATACA  
 2311 GGAGCAGATG ATACAGTATT AGAAGAAATG AGTTTGCCAG GAAGATGGAA ACCAAAAATG ATAGGGGGGA  
 2381 TCGGGGGCTT CATCAAGGTG AGGCAGTACG ACCAGATACT CATAGAAATC TGTGGACATA AAGCTATAGG  
 2451 TACAGTATTA GTAGGACCTA CACCTGTCAA CATAATTGGA AGAAATCTGT TGACCCAGAT CGGCTGCACC  
 2521 TTGAACTTCC CCATCAGCCC TATTGAGACG GTGCCCCTGA AGTTGAAGCC GGGGATGGAC GGCCCCAAGG  
 2591 TCAAGCAATG GCCATTGACG AAAGAGAAGA TCAAGGCCTT AGTCGAAATC TGTACAGAGA TGGAGAAGGA  
 2661 AGGGAAGATC AGCAAGATCG GGCCTGAGAA CCCCTACAAC ACTCCAGTCT TCGCAATCAA GAAGAAGGAC  
 2731 AGTACCAAGT GGAGAAAGCT GGTGGACTTC AGAGAGCTGA ACAAGAGAAC TCAGGACTTC TGGGAAGTTC  
 2801 AGCTGGGCAT CCCACATCCC GCTGGGTTGA AGAAGAAGAA GTCAGTGACA GTGCTGGATG TGGGTGATGC  
 2871 CTAATTCTCC GTTCCCTTGG ACGAGGACTT CAGGAAGTAC ACTGCCTTCA CGATACCTAG CATCAACAAC  
 2941 GAGACACCAG GCATCCGCTA CCAGTACAAC GTGCTGCCAC AGGGATGGAA GGCATACCA GCCATCTTTC  
 3011 AAAGCAGCAT GACCAAGATC CTGGAGCCCT TCCGCAAGCA AAACCAGAC ATCGTGATCT ATCAGTACAT  
 3081 GGACGACCTC TACGTAGGAA GTGACCTGGA GATCGGGCAG CACAGGACCA AGATCGAGGA GCTGAGACAG  
 3151 CATCTGTTGA GGTGGGGACT GACCACACCA GACAAGAAGC ACCAGAAGGA ACCTCCCTTC CTGTGGATGG  
 3221 GCTACGAACT GCATCCTGAC AAGTGACAG TGCAGCCCAT CGTGCTGCCT GAGAAGGACA GCTGGACTGT  
 3291 GAACGACATA CAGAAGCTCG TGGGCAAGTT GAACTGGGCA AGCCAGATCT ACCCAGGCAT CAAAGTTAGG  
 3361 CAGCTGTGCA AGCTGCTTCG AGGAACCAAG GCACTGACAG AAGTGATCCC ACTGACAGAG GAAGCAGAGC  
 3431 TAGAACTGGC AGAGAACCGA GAGATCCTGA AGGAGCCAGT ACATGGAGTG TACTACGACC CAAGCAAGGA  
 3501 CCTGATCGCA GAGATCCAGA AGCAGGGGCA AGGCCAATGG ACCTACCAAA TCTACCAGGA GCCCTTCAAG  
 3571 AACCTGAAGA CAGGCAAGTA CGCAAGGATG AGGGGTGCCC ACACCAACGA TGTGAAGCAG CTGACAGAGG  
 3641 CAGTGCAGAA GATCACCACA GAGAGCATCG TGATCTGGGG CAAGACTCCC AAGTTCAAGC TGCCCATACA  
 3711 GAAGGAGACA TGGGAGACAT GGTGGACCGA GTA CTGGCAA GCCACCTGGA TCCCTGAGTG GGAGTTCGTG

FIG. 9B

3781 AACACCCCTC CTTGGTGAA ACTGTGGTAT CAGCTGGAGA AGGAACCCAT CGTGGGAGCA GAGACCTTCT  
 3851 ACCTGGATGG GGCAGCCAAC AGGAGACCA AGCTGGGCAA GGCAGGCTAC GTGACCAACC GAGGACGACA  
 3921 GAAAGTGGTG ACCCTGACTG ACACCACCAA CCAGAAGACT GAGCTGCAAG CCATCTACCT AGCTCTGCAA  
 3991 GACAGCGGAC TGAAGTGAA CATCGTGACA GACTCACAGT ACGCACTGGG CATCATCCAA GCACAACCAG  
 4061 ACCAATCCGA GTCAGAGCTG GTGAACCAGA TCATCGAGCA GCTGATCAAG AAGGAGAAAG TGTACCTGGC  
 4131 ATGGGTACCA GCACACAAAG GAATTGGAGG AAATGAACAA GTAGATAAAT TAGTCAGTGC TGGGATCCGG  
 4201 AAGGTGCTGT TCCTGGACGG GATCGATAAG GCCCAAGATG AACATGAGAA GTACCACTCC AACTGGCGCG  
 4271 CTATGGCCAG CGACTTCAAC CTGCCACCTG TAGTAGCAAA AGAAATAGTA GCCAGCTGTG ATAAATGTCA  
 4341 GCTAAAAGGA GAAGCCATGC ATGGACAAGT AGACTGTAGT CCAGGAATAT GGCAGCTGGA CTGCACGCAC  
 4411 CTGGAGGGGA AGGTGATCCT GGTAGCAGT CATGTAGCCA GTGGATATAT AGAAGCAGAA GTTATCCCTG  
 4481 CTGAACTGG GCAGGAAACA GCATATTTTC TTTTAAATT AGCAGGAAGA TGGCCAGTAA AAACAATACA  
 4551 CACGGACAAC GGAAGCAACT TCACTGGTGC TACGGTTAAG GCCGCCTGTT GGTGGGCGGG AATCAAGCAG  
 4621 GAATTTGGAA TTCCCTACAA TCCCAATCG CAAGGAGTCG TGGAGAGCAT GAACAAGGAG CTGAAGAAGA  
 4691 TCATCGGACA AGTGAGGGAT CAGGCTGAGC ACCTGAAGAC AGCAGTGCAG ATGGCAGTGT TCATCCACAA  
 4761 CTTCAAAAGA AAAGGGGGGA TTGGGGGGTA CAGTGCAGGG GAAAGGATCG TGGACATCAT CGCCACCGAC  
 4831 ATCCAAACCA AGGAGCTGCA GAAGCAGATC ACCAAGATCC AGAACTCCG GGTGTACTAC CGCGACAGCC  
 4901 GCAACCCACT GTGGAAGGGA CCAGCAAAGC TCCTCTGGAA GGGAGAGGGG GCAGTGGTGA TCCAGGACAA  
 4971 CAGTGACATC AAAGTGGTGC CAAGGCGCAA GGCCAAGATC ATCCGCGACT ATGGAAAACA GATGCCAGGT  
 5041 GATGATTGTG TGGCAAGTAG ACAGGATGAG GATTAGAACC TGAAGAGCC TGGTGAAGCA CCATATGGCG  

  
 NheI (5117)  
 BstBI (5111)  
 5111 TTCGAAGCTA GCCTCGAGAT CCAGATCTGC TGTGCCTTCT AGTTGCCAGC CATCTGTTGT TTGCCCCTCC  
 5181 CCCGTGCCTT CTTGACCCT GGAAGGTGCC ACTCCCACTG TCCTTTCCTA ATAAATGAG GAAATTGCAT  
 5251 CGCATTGTCT GAGTAGGTGT CATTCTATTG TGGGGGGTGG GGTGGGGCAG CACAGCAAGG GGGAGGATTG  
 5321 GGAAGACAAT AGCAGGCATG CTGGGGATGC GGTGGGCTCT ATGCGTACCC AGGTGCTGAA GAATTGACCC  
 5391 GGTTCCTCCT GGGCCAGAAA GAAGCAGGCA CATCCCCTTC TCTCTGACAC ACCCTGTCCA CGCCCCTGGT  
 5461 TCTTAGTTCC AGCCCCACTC ATAGGACACT CATAGCTCAG GAGGGCTCCG CCTTCAATCC CACCCGCTAA  
 5531 AGTACTTGA GCGGTCTCTC CCTCCCTCAT CAGCCCACCA AACCAAACCT AGCCTCCAAG AGTGGGAAGA

FIG. 9C

5601 AATTAAAGCA AGATAGGCTA TTAAGTGCAG AGGGAGAGAA AATCCCTCCA ACATGTGAGG AAGTAATGAG

5671 AGAAATCATATA GAATTTCTTC CGCTTCCTCG CTCACTGACT CGCTGCGCTC GGTTCGTTCGG CTGCGGCGAG

5741 CCGGTATCAGC TCACTCAAAG GCGGTAATAC GGTATATCCAC AGAATCAGGG GATAACGCAG GAAAGAACAT

5811 GTGAGCAAAA GGGCAGCAAA AGGCCAGGAA CCGTAAAAAG GCCCGCTTGC TGGCGTTTTT CCATAGGCTC

5881 CGCCCCCTG ACAGGCATCA CAAAAATCGA CGCTCAAGTC ACAGGTGGCG AAACCCGACA GGAATAATAA

5951 GATACCAGGC GTTTCCCTCT GGAAGCTCCC TCCTGCTCCG ACCCTGCCGC TTACCGGATA

6021 CCTGTCCGCC TTTCTCCCTT CCGGAAGCGT GCGGCTTTCT CAATGCTCAC GCTGTAGGTA TCTCAGTTCC

6091 GTGTAGGTCG TTCGCTCCAA GCTCGGCTGT GTGCACGAAC CCCCCGTTC A GCGCGACCGC TCGCCCTTAT

6161 CCGGTAACTA TCCTCTTGAG TCCAACCCGG CTTATCGCCA CTGGCAGCAG CCACTGGTAA

6231 CAGGATTAGC AGAGCGAGGT ATGTAGCGCG TCTACACAG TTCTTGAAGT GGTGGCCTAA CTACCGGTAC

6301 ACTAGAAGCA CAGTATTGG TATCTGCGCT CTGCTGAAGC CAGTTACCTT CCGAAAAAGA GTTGCTAGCT

6371 CTTGATCCCG CAAACAAACC ACCGCTGGTA TATATGAGT AAAAGGATCT TCACCTAGAT CTTTAAAT

6441 AAAAAAAGGA TCTCAAGAAG ATCCTTTGAT CTTTTCTACG GGTCTGACG CTCAGTGGAA TTACCGCCAG

6511 CGTTAAGGGA TTTTGGTCAT GAGATTATCA AAAAGGATCT TCACCTAGAT CTTTAAAT

6581 GTTTTAAATC AATCTAAAGT ATATATGAGT AAAAGGATCT TCACCTAGAT CTTTAAAT

6651 ACCTATCTCA GCGATCTGTC TATTTCTGTT ATCCTTGCTC TGACAGTTAC CAATGCTTAA TCAGTGAGGC

6721 TCTGCCTCGT GAAGAAGGTG TTGCTGACTC ATCCATAGTT GCCTGACTCC GGGGGGGGGG GCGCGTGAGG

6791 AGGGAGCCAC GGTGTATGAG AGCTTTGTTG TAGGTGGACC TGAATCGCCC CATCATCCAG CCAGAAAGTG

6861 CGGAACGGTC TCGCTTGTCG GGAAGATGCG ACTTGGTGAT ACTTGGTGAT TTTGAACITT TGCTTTGCCA

6931 ACAAAGCCGC CGTCCCGTCA AGTCAGCGTA ATGCTCTGCC AGTGTTACAA GCAAAAGTTC GATTATTATCA

7001 AGAAAACTC ATCGAGCATC AAATGAAACT GCAATTTATT CATATCAGGA CCAATTAACC AATTCTGATT

271 PhePheGlu AspLeuMetL euHisPheG1 MetAspProA snAspIleG1 yTyrLysGln

7071 AAAAAGCCGT TTCTGTAATG AAGGAGAAAA CTCACCGAGG CAGTTCCATA GGATGGCAAG ATCTCGGTAT

248 PheLeuArgL ysGlnLeuSe rProSerPhe GluGlyLeuC ysAsnTrpLe uIleAlaLeu AspGlnTyrA

7141 CCGTCTGCGA TTCCGACTCG TCCAACATCA ATACAACCTA TTAATTTCCC CTGCTCAAAA ATAAGGTTAT

224 rgAspAlaI1 eGlyValArg GlyValAspI leCysGlyI1 eLeuLysGly GluAspPheI leLeuAsnAs

7211 CAAGTCAGAA ATCACCATGA GTGACGACTG AATCCGGTGA GAATGGCAAA AGCTTATGCA TTTCTTTCCA

201 pLeuSerPhe AspGlyHisT hrValValSe rAspProSer PheProLeuL euLysHisMe tGluLysTrp

7281 GACTTGTTC A CAGGCCAGC CATTACGCTC ATCATCAAAA TCACTCGCAT CAACCAAACC GTTATTTCATT

178 ValGlnGluV alProTrpG1 yAsnArgGlu AspAspPheA spSerAlaAs pValLeuGly AsnAsnMetA

7351 CGTGATTGCG CCTGACCGAG ACGAAATACG CGATCGCTGT TAAAAGGACA ATTACAAACA GGAATCGAAT

154 rgSerGlnAl aGlnAlaLeu ArgPheValA rgAspSerAs nPheProCys AsnCysValP roIleSerHi

7421 GCAACCGCG CAGGAACACT GCCACGCGAT CAACAATATT TTCACCTGAA TCAGGATATT CTCTAATAC

131 sLeuArgArg LeuPheValA laLeuAlaAs pValIleAsn GluGlySerA spProTyrG1 uGluLeuVal

7491 CTGGAATGCT GTTTTCCCGG CGATCGCAGT GGTGAGTAAC CATGCATCAT CAGGAGTACG GATAAAATGC

108 GlnPheAlaT hrLysGlyPr oIleAlaThr ThrLeuLeuT rpAlaAspAs pProThrArg ilePheHisL

7561 TTGATGGTGC GAAGAGGCAT AAATTCCGTC AGCCAGTTTA GTCTGACCAT CTCATCTGTA ACATCATTTGG

84 ysIleThrPr oLeuProMet PheGluThrL eTrpAsnLe uArgValMet GluAspThrV alAspAsnAl

7631 CAACGCTACC TTTGCCATGT TTCAGAAACA ACTCTGGCGC ATCGGGCTTC CCATACAATC CATAGATTGT

61 aValSerGly LysGlyHisL ysLeuPheLe uGluProAla AspProLysG lyTyrLeuAr gTyrIleThr

7701 CGCACCTGAT TGCCCGACAT TATCGCGAGC CCATTTATAC CCATATAAAT CAGCATCCAT GTTGAATTT

38 AlaGlySerG lnGlyValAs nAspArgAla TrpLysTyrG lyTyrLeuAs pAlaAspMet AsnSerAsnL

7771 AATCGCGGCC TCGAGCAAGA CGTTTCCCGT TGAATATGGC TCATAACACC CTTGTATTA CTGTTTATGT

14 euArgProAr gSerCysSer ThrGluArgG CATGATGATA TATTTTATC TTGTGCAATG TAACATCAGA GATTTTGAGA

7841 AAGCAGACAG TTTTATTGTT CCCCCCATTA TTTGAGCATT TATCAGGTTT ATTGTCTCAT GAGCGGATAC

7911 CACAACGTGG CTTTCCCCC CCCCCCATTA TTTGAGCATT TATCAGGTTT ATTGTCTCAT GAGCGGATAC

7981 ATATTGTAAT GTATTTAGAA AAATAAACAA ATAGGGGTTT CCGGCACATT TCCCCGAAA GTGCCACCTG

8051 ACGTCTAAGA AACCATTTAT ATCATGACAT TAACCTATAA AAATAGGCGT ATCAGGAGGC CCTTTCGTCT

8121 CGCGCGTTTC GGTGATGACG GTGAAAACCT CTGACACATG CAGCTCCCGG ACACGGTTCAC AGCTTGTCTG

8191 TAAGCGGATG CCGGAGCAGC ACAAGCCCGT CAGGGCGCGT CAGCGGGTGT TGGCGGGTGT CCGGGCTGGC

8261 TTAACATATGC GGCATCAGAC CAGATTGTAC GTAGAGTGCA CCATATGCGG TGTGAAATAC CGCACAGATG

8331 CGTAAGGAGA AAATACCGCA TCAGATTGGC TATTGG (SEQUENCE ID NO: 6)

FIG. 9D

1 TGGAAGGGCT AATTTGGTCC CAAAAAAGAC AAGAGATCCT TGATCTGTGG ATCTACCACA CACAAGGCTA  
 71 CTTCCCTGAT TGGCAGAACT ACACACCAGG GCCAGGGATC AGATATCCAC TGACCTTTGG ATGGTGCTTC  
 141 AAGTTAGTAC CAGTTGAACC AGAGCAAGTA GAAGAGGCCA AATAAGGAGA GAAGAACAGC TTGTTACACC  
 211 CTATGAGCCA GCATGGGATG GAGGACCCGG AGGGAGAAGT ATTAGTGTGG AAGTTTGACA GCCTCCTAGC  
 281 ATTTGTCAC ATGGCCCGAG AGCTGCATCC GGAGTACTAC AAAGACTGCT CACATCGAGC TTTCTACAAG  
 351 GGACTTTCCG CTGGGGACTT TCCAGGGAGG TGTGGCCTCG GCGGCACTGG GGAGTGGCGA GCCCTCAGAT  
 421 GCTACATATA AGCAGCTGCT TTTTGCCTGT ACTGGGTCTC TCTGGTTAGA CCAGATCTGA GCCTGGGAGC  
 491 TCTCTGGCTA ACTAGGGAAC CCACTGCTTA AGCCTCAATA AAGCTTGCCT TGAGTGCTCA AAGTAGTGTC  
 561 TGCCCGTCTG TTGTGTGACT CTGGTAACTA GAGATCCCTC AGACCCTTTT AGTCAGTGTG GAAAATCTCT  
 631 ACCAGTGGCG CCCGAACAGG GACTTGAAAG CGAAAGTAAA GCCAGAGGAG ATCTCTCGAC GCAGGACTCG  
 701 GCTTGCTGAA <sup>BssHII (711)</sup> GCGCGCacgg caagaggcga ggggcggcgC ctgACgagGa cgccaaaaat tttgactagc  
 771 ggaggctaga agggagagagC <sup>ClaI (830)</sup> TCGGTGCGAG AGCGTCAGTA TCAAGCGGGG GAGAATTAGA TCGATGGGAA  
 841 AAAATTTCGGT TAAGGCCAGG GGGAAAGAAA AAATATAAAT TAAAACATAT AGTATGGGCA AGCAGGGAGC  
 911 TAGAACGATT CGCAGTTAAT CCTGGCCTGT TAGAAACATC AGAAGGCTGT <sup>AccI (959)</sup> AGACAAATAC TGGGACAGCT  
 981 ACAACCATCC CTTGAGACAG GATCAGAAGA ACTTAGATCA TTATATAATA CAGTAGCAAC CCTCTATTGT  
 1051 GTGCATCAAA GGATACAGAT AAAAGACACC AAGGAAGCTT TAGACAAGAT AGAGGAAGAG CAAAACAAAA

FIG. 10A

1121 GTAAGAAAAA AGCACAGCAA GCAGCAGCTG ACACAGGACA CAGCAATCAG GTCAGCCAAA ATTACCCTAT  
 1191 AGTGCAGAAC ATCCAGGGGC AAATGGTACA TCAGGCCATA TCACCTAGAA CTTTAAACGA TAAGCTTGGG  
 1261 AGTTCCGCGT TACATAACTT ACGGTAAATG GCGCGCCTGG CTGACCGCCC AACGACCCCC GCCCATTGAC  
 1331 GTCAATAATG ACGTATGTTC CCATAGTAAC GCCAATAGGG ACTTTCCATT GACGTCAATG GGTGGAGTAT  
 1401 TTACGGTAAA CTGCCCCTT GGCAGTACAT CAAGTGTATC ATATGCCAAG TACGCCCCCT ATTGACGTCA  
 1471 ATGACGGTAA ATGCCCCGCC TGGCATTATG CCCAGTACAT GACCTTATGG GACTTTCCTA CTGGCAGTA  
 1541 CATCTACGTA TTAGTCATCG CTATTACCAT GGTGATGCGG TTTTGGCAGT ACATCAATGG GCGTGGATAG  
 1611 CGGTTTGA CT CACGGGGATT TCCAAGTCTC CACCCCAT TG ACGTCAATGG GAGTTTGTTT TGGCACCAAA  
 1681 ATCAACGGGA CTTTCCAAAA TGTCGTAACA ACTCCGCCCC ATTGACGCAA ATGGGCGGTA GCGGTGTACG  
 1751 GTGGGAGGTC TATATAAGCA GAGCTCGTTT AGTGAACCGT CAGATCGCCT GGAGACGCCA TCCACGCTGT  
 1821 TTTGACCTCC ATAGAAGACA CCGACTCTAG AGgatccATC TAAGTAAGCT TGGCATTCCG G TACTGTTGG  
 1891 TAAAATGGAA GACGCCAAAA ACATAAAGAA AGGCCCGGCG CCATTCTATC CTCTAGAGGA TGGAACCGCT  
 1961 GGAGAGCAAC TGCATAAGGC TATGAAGAGA TACGCCCTGG TTCCTGGAAC AATTGCTTTT ACAGATGCAC  
 2031 ATATCGAGGT GAACATCAGC TACGCGGAAT ACTTCGAAAT GTCCGTTCCG TTGGCAGAAG CTATGAAACG  
 2101 ATATGGGCTG AATACAAATC ACAGAATCGT CGTATGCAGT GAAAAC TCTC TTCAATTCTT TATGCCGGTG  
 2171 TTGGGCCCCG TATTTATCGG AGTTGCAGTT GCGCCCGCGA ACGACATTTA TAATGAACGT GAATTGCTCA  
 2241 ACAGTATGAA CATTTCGCAG CCTACCGTAG TGTGTTGTTT CAAAAAGGGG TTGCAAAAAA TTTTGAACGT  
 2311 GCAAAAAAAA TTACCAATAA TCCAGAAAAT TATTATCATG GATTCTAAAA CGGATTACCA GGGATTTCAG

FIG. 10B

2381 TCGATGTACA CGTTCGTAC ATCTCATCTA CCTCCCGGT TTAATGAATA CGATTTTGTA CCAGAGTCCT  
 2451 TTGATCGTGA CAAAACAATT GCACTGATAA TGAATTCCTC TGGATCTACT GGGTACCTA AGGGTGTGGC  
 2521 CCTTCCGCAT AGAACTGCCT GCGTCAGATT CTCGCATGCC AGAGATCCTA TTTTGGCAA TCAAATCATT  
 2591 CCGGATACTG CGATTTTAAG TGTGTGTTCCA TTCCATCAGG GTTTTGAAT GTTTACTACA CTCGGATATT  
 2661 TGATATGTGG ATTTGAGTGC GTCTTAATGT ATAGATTIGA AGAAGAGCTG TTTTACGAT CCCTTCAGGA  
 2731 TTACAAAATT CAAAGTGCGT TGCTAGTACC AACCTATTT TCATTCTTCG CAAAAGCAC TCTGATTGAC  
 2801 AAATACGATT TATCTAATTT ACACGAAATT GCTTCTGGGG GCGCACCTCT TTCGAAAGAA GTCGGCGAAG  
 2871 CGGTTGCAAA ACGCTTCCAT CTTCCAGGGA TACGACAAGG ATATGGGCTC ACTGAGACTA CATCAGCTAT  
 2941 TCTGATTACA CCCGAGGGGG ATGATAAACC GGGCGCGGTC GGTAAAGTTG TTCCATTTTT TGAAGCGAAG  
 3011 GTTGTGGATC TGGATACCGG GAAAACGCTG GCGTTAATC AGAGAGGCGA ATTATGTGTC AGAGGACCTA  
 3081 TGATTATGTC CGGTTATGTA AACAAATCCGG AAGCGACCAA CGCCTTGATT GACAAGGATG GATGGCTACA  
 3151 TTCTGGAGAC ATAGCTTACT GGGACGAAGA CGAACACTTC TTCATAGTTG ACCGCTTGAA GTCTTTAATT  
 3221 AAATACAAAG GATATCAGGT GGGCCCCGCT GAATTGGAAT CGATATTGTT ACAACACCCC AACATCTTCG  
 3291 ACCCGGGCGT GGCAGGTCTT CCCGACGATG ACGCCGGTGA ACTTCCCGCC GCCGTTGTTG TTTGGAGCA  
 3361 CGGAAAGACG ATGACGGAAG AAGAGATCGT GGATTACGTC GCCAGTCAAG TAACAACCGC GAAAAAGTTG  
 3431 CGCGGAGGAG TTGTGTTTGT GGACGAAGTA CCGAAAGGTC TTACCGGAAA ACTCGACGCA AGAAAAATCA  
 3501 GAGAGATCCT CATAAAGGCC AAGAAGGCCG GAAAGTCCAA ATTGTAACTC GAGGGGGGGC CCGGTACCTT  
 ClaI (3259)  
 XhoI (3548)      ApaI (3557)  
 KpnI (3563)

FIG. 10C

3571 TAAGACCAAT GACTTACAAG GCAGCTGTAG ATCTTAGCCA CTTTTTAAAA GAAAAGGGGG GACTGGAAGG  
 3641 GCTAATTAC TCCCAAAGAA GACAAGATAT CCTTGATCTG TGGATCTACC ACACACAAGG CTACTTCCCT  
 3711 GATTGGCAGA ACTACACACC AGGGCCAGGG GTCAGATATC CACTGACCTT TGGATGGTGC TACAAGCTAG  
 3781 TACCAGTTGA GCCAGATAAG GTAGAAGAGG CCAATAAAGG AGAGAACACC AGCTTGTTAC ACCCTGTGAG  
 3851 CCTGCATGGA ATGGATGACC CTGAGAGAGA AGTGTTAGAG TGGAGGTTTG ACAGCCGCCT AGCATTTCAT  
 3921 CACGTGGCCC GAGAGCTGCA TCCGGAGTAC TTCAAGAACT GCTGACATCG AGCTTGCTAC AAGGGACTTT  
 3991 CCGCTGGGGA CTTTCCAGGG AGGCGTGGCC TGGGCGGGAC TGGGAGTGG CGAGCCCTCA GATGCTGCAT  
 4061 ATAAGCAGCT GCTTTTGGC TGTACTGGGT CTCTCTGGTT AGACCAGATC TGAGCCTGGG AGCTCTCTGG  
 4131 CTAAGTAGGG AACCCACTGC TTAAGCCTCA ATAAAGCTTG CCTTGAGTGC TTCAAGTAGT GTGTGCCCCG  
 4201 CTGTTGTGTG ACTCTGGTAA CTAGAGATCC CTCAGACCCT TTAGTCACT GTGGAAAATC TCTAGCACCC  
 4271 CCCAGGAGGT AGAGGTTGCA GTGAGCCAAG ATCGCGCCAC TGCATTCCAG CCTGGGCAAG AAAACAAGAC  
 4341 TGTCTAAAAA AATAATAATA AGTTAAGGGT ATTAATATATA TTTATACATG GAGGTCATAA AAATATATAT  
 4411 ATTTGGGCTG GCGCGAGTGG CTCACACCTG CGCCC GGCCC TTTGGGAGGC CGAGGCAGGT GGATCACCTG  
 4481 AGTTTGGGAG TTCCAGACCA GCCTGACCAA CATGGAGAAA CCCCTTCTCT GTGTATTTTT ATGAGATTTT  
 4551 ATTTTATGTG TATTTTATTC ACAGGTATTT CTGGA AAACT GAAACTGTTT TTCCTCTACT CTGATACCAC  
 4621 AAGAATCATC AGCAGAGAGG AAGACTTCTG TGATCAAATG TGGTGGGAGA GGGAGGTTTT CACCAGCACA  
 4691 TGAGCAGTCA GTTCTGCCGC AGACTCGGCG GGTGTCCTTC GGTTCACTTC CAACACCCGC TGCCTGGAGA  
 4761 GAGGTCAGAC CACAGGGTGA GGGCTCAGTC CCCAAGACAT AAACACCCAA GACATAAACA CCCAACAGGT  
 4831 CCACCCCGCC TGCTGCCAG GCAGAGCCGA TTCACCAAGA CGGGAATTAG GATAGAGAAA GAGTAACTCA  
 4901 CACAGAGCCG GCTGTGCGGG AGAACGGAGT TCTATTATGA CTCAAATCAG TCTCCCAAG CATTCCGGGA  
 4971 TCAGAGTTTT TAAGGATAAC TTAGTGTGTA CGGGGCCAGT GAGTTGGAGA TGAAGCGTA GGGAGTCGAA  
 5041 GGTGTCTTTT TCGCGCCAGT CAGTTCCTGG GTGGGGGCCA CAAGATCGGA TGAGCCAGTT TATCAATCCG  
 5111 GGGGTGCCAG CTGATCCATG GAGTGCAGGG TCTGCAAAAT ATCTCAAGCA CTGATTGATC TTAGGTTTTA  
 5181 CAATAGTGAT GTTACCCAG GAACAATTG GGAAGGTCA GAATCTTGTA GCCTGTAGCT GCATGACTCC  
 5251 TAAACCATAA TTTCTTTTTT GTTTTTTTTT TTTTATTTTT GAGACAGGGT CTCACTCTGT CACCTAGGCT  
 5321 GGAGTGCAGT GGTGCAATCA CAGTCACTG CAGCCCTAG AGCGCGCCGC ACCGCGGTGG AGCTCCAATT  
 5391 CGCCCTATAG TGAGTCGTAT TACAATTCAC TGGCCGTCTG TTTACAACGT CGTGACTGGG AAAACCCCTGG  
 5461 CGTTACCCAA CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC  
 5531 ACCGATCGCC CTTCCCAACA GTTGCGCAGC CTGAATGGCG AATGGCGCGA AATTGTAAAC GTTAATATTI  
 5601 TGTTAAAAAT CGCGTTAAAT TTTTGTAAAA TCAGCTCATT TTTTAACCAA TAGGCCGAAA TCGGCAAAAT  
 5671 CCCTTATAAA TCAAAAGAAT AGACCGAGAT AGGGTTGAGT GTTGTTCCAG TTTGGAACAA GAGTCCACTA  
 5741 TTAAAGAACG TGGACTCCAA CGTCAAAGCG CGAAAAACCG TCTATCAGGG CGATGGCCCA CTACGTGAAC  
 5811 CATCACCTTA ATCAAGTTTT TTGGGGTCTGA GGTGCCGTAA AGCACTAAAT CGGAACCTTA AAGGGAGCCC  
 5881 CCGATTAGA GCTTGACGGG GAAAGCCGGC GAACGTGGCG AGAAAGGAAG GGAAGAAAGC GAAAGGAGCG

FIG. 10D

```

5951 GCGCGTAGGG CGCTGGCAAG TGTAGCGGTC ACGCTGCGCG TAACCACCAC ACCCGCCGCG CTTAATGCGC
6021 CGCTACAGGG CGCGTCCCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTC TTTATTTTTC
6091 TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA
6161 GGAACAGTAT GAGTATTCAA CATTTCGGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT GCCTTCCTGT
6231 TTTTGCTCAC CCAGAAACGC TGGTGAAGT AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTAC
6301 ATCGAACTGG ATCTCAACAG CGGTAAGATC CTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA
6371 GCACTTTTAA AGTTCTGCTA TGTGGCGCGG TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTCC
6441 CCGCATACAC TATTCTCAGA ATGACTTGCT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC
6511 ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TCGCGCCAAC TTACTTCTGA
6581 CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA CTCGCCCTGA
6651 TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA CCACGATGCC TGTAGCAATG
6721 GCAACAACGT TGCGCAAACT ATTAAGTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT
6791 GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGGTGGT TTATTGCTGA
6861 TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GCGCAGATGG TAAGCCCTCC
6931 CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA
7001 TAGGTGCCCTC ACTGATTAA GATTGGTAAC TGTGAGACCA AGTTTACTCA TATATACTTT AGATTGATTT
7071 AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA ACTTCATGAC CAAAATCCCT
7141 TAACGTGACT TTTTCGTTCCA CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT
7211 TTTTCTGCG CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA
7281 TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATAACAAA TACTGTCTTT
7351 CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
7421 TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA GACGATAGTT
7491 ACCGGATAAG GCGCAGCGGT CCGGCTGAAC GGGGGGTTCTG TGCACACAGC CCAGCTTGA GCGAAGCAGC
7561 TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG
7631 ACAGGTATCC GGTAAGCGGC AGGGTCGGA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCGTG
7701 GTATCTTTAT AGTCCTGTCT GGTTCGCCA CCTCTGACTT GAGCGTGGAT TTTTGTGATG CTCGTCAGGG
7771 GGGCGGAGCC TATGGAAAAA CGCCAGCAAC CGGGCCTTTT TACGGTTTCT GGCCTTTTGC TGGCCTTTTG
7841 CTCACATGTT CTTTCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG AGTGAGCTGA
7911 TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCCAATA
7981 CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA CGACAGGTTT CCCGACTGGA
8051 AAGCGGGCAG TGAGCGCAAC GCAATTAATG TGAGTTAGCT CACTCATTAG GCACCCGAGG CTTTACACTT
8121 TATGCTTCCG GCTCGTATGT TGTGTGGAAT TGTGAGCGGA TAACAATTTT ACACAGGAAA CAGCTATGAC
8191 CATGATTACG CCAAGCTCGG AATTAACCCT CACTAAAGGG AACAAAAGCT GCTGCAGGGT CCCTAACTGC
8261 CAAGCCCCAC AGTGTCGCTT GAGGCTGCCC CTTCCTTCTA GCGGCTGCCC CCACTCGGCT TTGCTTTCCC
8331 TAGTTTCAGT TACTTGGCTT CAGCCAAGGT CTGAAACTAG GTGCGCACAG AGCGGTAAGA CTGCGAGAGA
8401 AAGAGACCAG CTTTACAGGG GGTTTATCAC AGTGCACCCT GACAGTCGTC AGCCTCACAG GGGGTTTATC
8471 ACATTGCACC CTGACAGTCG TCAGCCTCAC AGGGGGTTTA TCACAGTGCA CCCTTACAAT CATTCCATTT
8541 GATTACAAT TTTTTTAGTC TCTACTGTGC CTAACCTGTA AGTTAAATTT GATCAGAGGT GTGTTCCAG
8611 AGGGGAAAAAC AGTATATACA GGGTTTCAGTA CTATCGCATT TCAGGCCTCC ACCTGGGTCT TGGAAATGTT
8681 CCCCCGAGG GTGATGACTA CCTCAGTTGC ATCTCCACAG GTCACAGTGA CACAAGATAA CCAAGACACC
8751 TCCCAAGGCT ACCACAATGG GCCGCCCTCC ACGTGCACAT GGCCGGAGGA ACTGCCATGT CCGAGGTGCA
8821 AGCACACCTG CGCATCAGAG TCCTTGGTGT GGAGGGAGGG ACCAGCGCAG CTTCCAGCCA TCCACCTGAT
8891 GAACAGAACC TAGGGAAAGC CCCAGTTCTA CTTACACCAG GAAAGGC (SEQUENCE ID NO: 8)

```

FIG. 10E

1 TGGAAGGGCT AATTGGTCC CAAAAAAGAC AAGAGATCCT TGATCTGTGG ATCTACCACA CACAAGGCTA  
 71 CTTCCCTGAT TGGCAGAACT ACACACCAGG GCCAGGGATC AGATATCCAC TGACCTTTGG ATGGTGCTTC  
 141 AAGTTAGTAC CAGTTGAACC AGAGCAAGTA GAAGAGGCCA AATAAGGAGA GAAGAACAGC TTGTTACACC  
 211 CTATGAGCCA GCATGGGATC GAGGACCCGG AGGGAGAAGT ATTAGTGTGG AAGTTTGACA GCCTCCTAGC  
 281 ATTCGTCAC ATGGCCCGAG AGCTGCATCC GGAGTACTAC AAAGACTGCT GACATCGAGC TTTCTACAAG  
 351 GGACTTTCGG CTGGGGACTT TCCAGGGAGG TGTGGCCTGG GCGGGACTGG GGAGTGGCGA GCCCTCAGAT  
 421 GCTACATATA AGCAGCTGCT TTTTGCCTGT ACTGGGTCTC TCTGGTTAGA CCAGATCTGA GCCTGGGAGC  
 491 TCTCTGGCTA ACTAGGGAAC CCACTGCTTA AGCCTCAATA AAGCTTGCCT TGAGTGCTCA AAGTAGTGTG  
 561 TGCCCGTCTG TTGTGTGACT CTGGTAACTA GAGATCCCTC AGACCCTTTT ACTCAGTGTG GAAAATCTCT  
 631 AGCAGTGGCG CCCGAACAGG GACTTGAAAG CGAAAGTAAA GCCAGAGGAG ATCTCTCGAC GCAGGACTCG  
 701 GCTTGCTGAA <sup>BssHII (711)</sup> GCGCGCacgg caagaggcga ggggcggcgC ctgACgagGa cgccaaaaat tttgactagc  
 771 ggaggctaga aggagagagC TCGGTCCGAG AGCGTCAGTA TTAAGCGGGG GAGAATTAGA TCGATGGGAA <sup>ClaI (830)</sup>  
 841 AAAATTCGGT TAAGGCCAGG GGGAAAGAAG AAGTACAAGC TAAAGCACAT CGTATGGGCA AGCAGGGAGC  
 911 TAGAACGATT CGCAGTTAAT CCTGGCCTGT TAGAAACATC AGAAGGCTGT AGACAAATAC TGGGACAGCT <sup>AccI (959)</sup>  
 981 ACAACCATCC CTCAGACAG GATCAGAGGA GCTTCGATCA CTATACAACA CAGTAGCAAC CCTCTATTGT  
 1051 GTGCACCAGC GGATCGAGAT CAAGGACACC AAGGAAGCTT TAGACAAGAT AGAGCAAGAG CAAAACAAGT  
 1121 CCAAGAAGAA GGCCAGCAG GCAGCAGCTG ACACAGGACA CAGCAATCAG GTCAGCCAAA ATTACCCTAT

FIG. 11A

```

1191 AGTGCAGAAC ATCCAGGGGC AAATGGTACA TCAGGCCATA TCACCTAGAA CTTTAAACGA TAAGCTTGGG
1261 AGTTCCGCGT TACATAACTT ACGGTAAATG GCGCGCCTGG CTGACCGCCC AACGACCCCC GCCCATTGAC
1331 GTCAATAATG ACGTATGTTC CCATAGTAAC GCCAATAGGG ACTTTCCATT GACGTCAATG GGTGGAGTAT
1401 TTACGGTAAA CTGCCCACCTT GGCAGTACAT CAAGTGATC ATATGCCAAG TACGCCCCCT ATTGACGTCA
1471 ATGACGGTAA ATGGCCCCGC TGGCATTATG CCCAGTACAT GACCTTATGG GACTTTCCTA CTTGGCAGTA
1541 CATCTACGTA TTAGTCATCG CTATTACCAT GGTGATGCGG TTTTGGCAGT ACATCAATGG GCGTGGATAG
1611 CGGTTTGA CT CACGGGGATT TCCAAGTCTC CACCCCATTG ACGTCAATGG GAGTTTGT TT TGGCACCAAA
1681 ATCAACGGGA CTTTCCAAAA TGTCGTAACA ACTCCGCCCC ATTGACGCAA ATGGGCGGTA GCGGTGTACG
1751 GTGGGAGGTC TATATAAGCA GAGCTCGTTT AGTGAACCGT CAGATCGCCT GGACACGCCA TCCACGCTGT
1821 TTTGACCTCC ATAGAAGACA CCGACTCTAG AGgatccATC TAAGTAAGCT TGGCATTCCG GTACTGTTGG
1891 TAAAATGGAA GACGCCAAAA ACATAAAGAA AGGCCCGGCG CCATTCTATC CTCTAGAGGA TGAACCGCT
1961 GGAGAGCAAC TGCATAAGGC TATGAAGAGA TACGCCCTGG TTCCTGGAAC AATTGCTTTT ACAGATGCAC
2031 ATATCGAGGT GAACATCAGC TACGCGGAAT ACTTCGAAAT GTCCGTTCCG TTGGCAGAAG CTATGAAACG
2101 ATATGGGCTG AATACAAATC ACAGAATCGT CGTATGCAGT GAAACTCTC TTCAATTCTT TATGCCGGTG
2171 TTGGCGCGCT TATTTATCGG AGTTGCAGTT GCGCCCGCGA ACGACATTTA TAATGAACGT GAATTGCTCA
2241 ACAGTATGAA CATTTCGCAG CCTACCGTAG TGTITGTTTC CAAAAGGGG TTGCAAAAAA TTTTGAACGT
2311 GCAAAAAAAA TTACCAATAA TCCAGAAAAT TATTATCATG GATTCTAAAA CGGATTACCA GCGATTTCAG
2381 TCGATGTACA CGTTCGTAC ATCTCATCTA CCTCCCGGTT TTAATGAATA CGATTTTGTA CCAGAGTCCT

```

FIG. 11B

2451 TTGATCGTGA CAAAACAATT GCACTGATAA TGAATTCCTC TGGATCTACT GGGTTACCTA AGGGTGTGGC

2521 CCTTCCGCAT AGAACTGCCT GCGTCAGATT CTCGCATGCC AGAGATCCTA TTTTGGCAA TCAAATCATT

2591 CCGGATACTG CGATTTTAAG TGTTGTTCCA TTCCATCAGG GTTTTGAAT GTTTACTACA CTCGGATATT

2661 TGATATGTGG ATTTGAGTC GTCTTAATGT ATAGATTTGA AGAAGAGCTG TTTTACGAT CCCTTCAGGA

2731 TTACAAAATT CAAAGTGGT TGCTAGTACC AACCTATTT TCATTCTTCG CCAAAGCAC TCTGATTGAC

2801 AAATACGATT TATCTAATTT ACACGAAATT GCTTCTGGGG GCGCACCTCT TTCGAAAGAA GTCGGGGAAG

2871 CGGTTGCAAA ACGCTTCCAT CTTCCAGGGA TACGACAAGG ATATGGGCTC ACTGAGACTA CATCAGCTAT

2941 TCTGATTACA CCCGAGGGGG ATGATAAACC GGGCGCGGTC GGTAAAGTTG TTCCATTTTT TGAAGCGAAG

3011 GTTGTGGATC TGGATACCGG GAAAACGCTG GCGGTTAATC AGAGAGGCGA ATTATGTGTC AGAGGACCTA

3081 TGATTATGTC CGGTTATGTA AACAATCCGG AAGCGACCAA CGCCTTGATT GACAAGGATG GATGGCTACA

3151 TTCTGGAGAC ATAGCTTACT GGGACGAAGA CGAACACTTC TTCATAGTTG ACCGCTTGAA GTCTTTAATT

3221 AAATACAAAG GATATCAGGT GGCCCCGCT GAATTGGAAT CGATATTGTT ACAACACCCC AACATCTTCG

3291 ACGCGGGCGT GGCAGGTCTT CCCGACGATG ACGCCGGTGA ACTTCCCGCC GCCGTTGTTG TTTTGGAGCA

3361 CGGAAAGACG ATGACCGAAA AAGAGATCGT GGATTACGTC GCCAGTCAAG TAACAACCGC GAAAAAGTTG

3431 CGCGGAGGAG TTGTGTTTGT GGACGAAGTA CCGAAAGGTC TTACCGGAAA ACTCGACGCA AGAAAAATCA

3501 GAGAGATCCT CATAAAGGCC AAGAAGGGCG GAAAGTCCAA ATTGTAAcTC GAGGGGGGGC CCGGTACCTT

3571 TAAGACCAAT GACTTACAAG GCAGCTGTAG ATCTTAGCCA CTTTTTAAAA GAAAAGGGGG GACTGGAAGG

Clal (3259)

XhoI (3548)      ApaI (3557)      KpnI (3563)

FIG. 11C

3641 GCTAATTCAC TCCCAAAGAA GACAAGATAT CCTTGATCTG TGGATCTACC ACACACAAGG CTACTTCCCT  
 3711 GATTGGCAGA ACTACACACC AGGGCCAGGG GTCAGATATC CACTGACCTT TGGATGGTGC TACAAGCTAG  
 3781 TACCAGTTGA GCCAGATAAG GTAGAAGAGG CCAATAAAGG AGAGAACACC AGCTTGTTAC ACCCTGTGAG  
 3851 CCTGCATGGA ATGGATGACC CTGAGAGAGA AGTGTTAGAG TGGAGGTTTG ACAGCCGCCT AGCATTTCAT  
 3921 CACGTGGCCC GAGAGCTGCA TCCGGAGTAC TTCAAGAACT GCTGACATCG AGCTTGCTAC AAGGGACTTT  
 3991 CCGCTGGGGA CTTTCCAGGG AGGCGTGGCC TGGGCGGGAC TGGGAGTGG CGAGCCCTCA GATGCTGCAT  
 4061 ATAAGCAGCT GCTTTTGGCC TGTACTGGGT CTCTCTGGTT AGACCAGATC TGAGCCTGGG AGCTCTCTGG  
 4131 CTAAGTAGGG AACCCACTGC TTAAGCCTCA ATAAAGCTTG CCTTGAGTGC TTCAAGTAGT GTGTGCCCGT  
 4201 CTGTTGTGTG ACTCTGGTAA CTAGAGATCC CTCAGACCCT TTTAGTCAGT GTGGAAAATC TCTAGCACCC  
 4271 CCCAGGAGGT AGAGGTTGCA GTGAGCCAAG ATCGCGCCAC TGCATTCCAG CCTGGGCAAG AAAACAAGAC  
 4341 TGCTAAAAAT AATAATAATA AGTTAAGGGT ATTAATATATA TTTATACATG GAGGTCATAA AAATATATAT  
 4411 ATTTGGGCTG GGCGCAGTGG CTCACACCTG CGCCCGGCCC TTTGGGAGGC CGAGGCAGGT GGATCACCTG  
 4481 AGTTTGGGAG TTCCAGACCA GCCTGACCAA CATGGAGAAA CCCCTTCTCT GTGTATTTTT AGTAGATTTT  
 4551 ATTTTATGTG TATTTIATTC ACAGGTATTT CTGGAAAACT GAAACTGTTT TTCCTCTACT CTGATACCAC  
 4621 AAGAATCATC AGCACAGAGG AAGACTTCTG TGATCAAAATG TGGTGGGAGA GGGAGGTTTT CACCAGCACA  
 4691 TGAGCAGTCA GTTCTGCCGC AGACTCGCGG GGTGTCCTTC GGTTCAGTTC CAACACCGCC TGCCCTGGAGA  
 4761 GAGGTCAGAC CACAGGCTGA GGGCTCAGTC CCCAAGACAT AAACACCCAA GACATAAACA CCCAACAGGT  
 4831 CCACCCCGCC TGCTGCCCAG GCAGAGCCGA TTCACCAAGA CGGGAATTAG GATAGAGAAA GAGTAAAGTCA  
 4901 CACAGAGCCG GCTGTGCGGG AGAACGGAGT TCTATTATGA CTCAAATCAG TCTCCCAAG CATTCGGGGA  
 4971 TCAGAGTTTT TAAGGATAAC TTAGTGTGTA GGGGCCAGT GAGTTGGAGA TGAAAGCGTA GGGAGTCGAA  
 5041 GCTGTCTTTT TGCGCCGAGT CAGTTCCTCG GTGGGGGCCA CAAGATCGGA TGAGCCAGTT TATCAATCCG  
 5111 GGGGTGCCAG CTGATCCATG GAGTGCAGGG TCTGCAAAAT ATCTCAAGCA CTGATTGATC TTAGGTTTTA  
 5181 CAATAGTGAT GTTACCCAG GAACAATTG GGGAAAGTCA GAATCTTGTA GCCGTAGCT GCATGACTCC  
 5251 TAAACCATAA TTTCTTTTT GTTTTTTTTT TTTTATTTTT GAGACAGGGT CTCACTCTGT CACCTAGGCT  
 5321 GGAGTGCAGT GGTGCAATCA CAGCTCACTG CAGCCCCTAG AGCGCCGCC ACCCGCGTGG AGCTCCAATT  
 5391 CGCCCTATAG TGAGTCGTAT TACAATTCAC TGGCCGTCGT TTTACAACGT CGTGAAGTGG AAAACCCCTGG  
 5461 CGTTACCCAA CTTAATCGCC TTGCAGCACA TCCCCCTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC  
 5531 ACCGATCGCC CTTCCCAACA GTTGCAGCAG CTGAATGGCG AATGGCGCGA AATTGTAAAC GTTAATATTT  
 5601 TGTTAAAAAT CGCGTTAAAT TTTTGTTAAA TCAGCTCATT TTTTAACCAA TAGGCCGAAA TCGGCAAAAT  
 5671 CCCTTATAAA TCAAAAGAAT AGACCGAGAT AGGTTGAGT GTTGTTCAG TTTGGAACAA GAGTCCACTA  
 5741 TTAAGAAGCG TGGACTCCAA CGTCAAAGGG CGAAAAACCG TCTATCAGGG CGATGGCCCA CTACGTGAAC  
 5811 CATCACCTTA ATCAAGTTTT TTGGGGTCGA GGTGCCGTAA AGCACTAAAT CGGAACCTTA AAGGGAGCCC  
 5881 CCGATTTAGA GCTTGACGGG GAAAGCCGGC GAACGTGGCG AGAAAGGAA GGAAGAAAGC GAAAGGAGCG  
 5951 GCGGCTAGGG CGCTGGCAAG TGTAGCCGTC ACCGTGCGCG TAACCACCAC ACCCGCCCGC CTTAATCGCG  
 6021 CGCTACAGGG CGCGTCCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTT  
 6091 TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA

FIG. 11D

```

6161 GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT GCCTTCCTGT
6231 TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGCTTAC
6301 ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA
6371 GCACTTTTAA AGTTCTGCTA TGTGGCGCGG TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTCC
6441 CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC
6511 ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAAGCATGA GTGATAACAC TCGCGCCAAC TTACTTCTGA
6581 CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTTGA CAACATGGGG GATCATGTAA CTCGCCTTGA
6651 TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAAACGAC GAGCGTGACA CCACGATGCC TGTAGCAATG
6721 GCAACAACGT TCGCGAAACT ATTAAGTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT
6791 GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC GCGCCTTCCG GCTGGCTGGT TTATTGCTGA
6861 TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CCGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC
6931 CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA
7001 TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTACAGCCA AGTTTACTCA TATATACTTT AGATTGATTT
7071 AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT
7141 TAACGTGAGT TTTCTTCCA CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT
7211 TTTTTCTGCG CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA
7281 TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTCTTT
7351 CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAAGTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
7421 TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCTGT TCTTACCGGG TTGGACTCAA GACGATAGTT
7491 ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGTTTCG TGCACACAGC CCAGCTTGGG GCGAACGACC
7561 TACACCGAAG TGAGATACCT ACAGCGTGAG CTATGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG
7631 ACAGGTATCC GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCAGGGG GAAACGCCTG
7701 GTATCTTTAT AGTCCTGTCT GGTTCGCCA CCTCTGACTT GAGCGTCCAT TTTTGTGATG CTCGTCAGGG
7771 GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG
7841 CTCACATGTT CTTTCTTCCG TTATCCCTTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG AGTGAGCTGA
7911 TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCCAATA
7981 CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA CGACAGGTTT CCGGACTGGA
8051 AAGCGGGCAG TGAGCGCAAC GCAATTAATG TGAGTTAGCT CACTCATTAG GCACCCCAAG CTTTACACTT
8121 TATGCTTCCG GCTCGTATGT TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC
8191 CATGATTACC CCAAGCTCGG AATTAACCTT CACTAAAGGG AAAAAAGCT GCTGCAGGGT CCCTAACTGC
8261 CAAGCCCCAC AGTGTGCCCT GAGGCTGCCC CTTCTTCTA CCGGCTGCCC CCACTCGGCT TTGCTTTCCC
8331 TAGTTTCAGT TACTTGCGTT CAGCCAAGGT CTGAACTAG GTGCGCACAG AGCGGTAAGA CTGCGAGAGA
8401 AAGAGACCAG CTTTACAGGG GGTTTATCAC AGTGCACCCT GACAGTCGTC AGCCTCACAG GGGGTTTATC
8471 ACATTGCACC CTGACAGTCG TCAGCCTCAC AGGGGGTTTA TCACAGTGCA CCCTTACAAT CATTCATTT
8541 GATTACAAT TTTTTTAGTC TCTACTGTCC TAACTTTGTA AGTTAAATTT GATCAGAGGT GTGTTCCAG
8611 AGGGGAAAAA AGTATATACA GGGTTCAGTA CTATCGCATT TCAGGCCTCC ACCTGGGTCT TGGAAATGTG
8681 CCCCCAGGG GTGATGACTA CCTCAGTTGG ATCTCCACAG GTCACAGTGA CACAAGATAA CCAAGACACC
8751 TCCCAAGGCT ACCACAATGG GCGGCCCTCC ACGTGACAT GGCCGGAGGA ACTGCCATGT CGGAGGTGCA
8821 AGCACACCTG CGCATCAGAG TCCTTGGTGT GGAGGGAGGG ACCAGCGCAG CTTCCAGCCA TCCACCTGAT
8891 GAACAGAACC TAGGGAAAGC CCCAGTTCTA CTTACACCAG GAAAGGC (SEQUENCE ID NO: 9)

```

FIG. 11E

mBCwCN frag	-----C- --AC--G--	-----
m2BCwCN frag	-----C- --G--G--	-----
BC/HXB2	-----	-----
BC/NL43	-----	-----
#1	.....	.....
	CGCGCACGGC AAGAGGCGAG GGGCGGCGAC TGGTGAGTAC GCCAAAAATT	
mBCwCN frag	-----C- C-----	-----
m2BCwCN frag	-----	-----
BC/HXB2	-----T--	-----
BC/NL43	-----	-----G----
#51	.....	.....
	TTGACTAGCG GAGGCTAGAA GGAGAGAGAT <u>GGGTGCGAGA</u> GCGTCAGTAT	
mBCwCN frag	-----	-----
m2BCwCN frag	-----	-----
BC/HXB2	-----	-----
BC/NL43	-----AA-----	-----
#101	.....	.....
	TAAGCGGGGG AGAATTAGAT CG	

FIG. 12

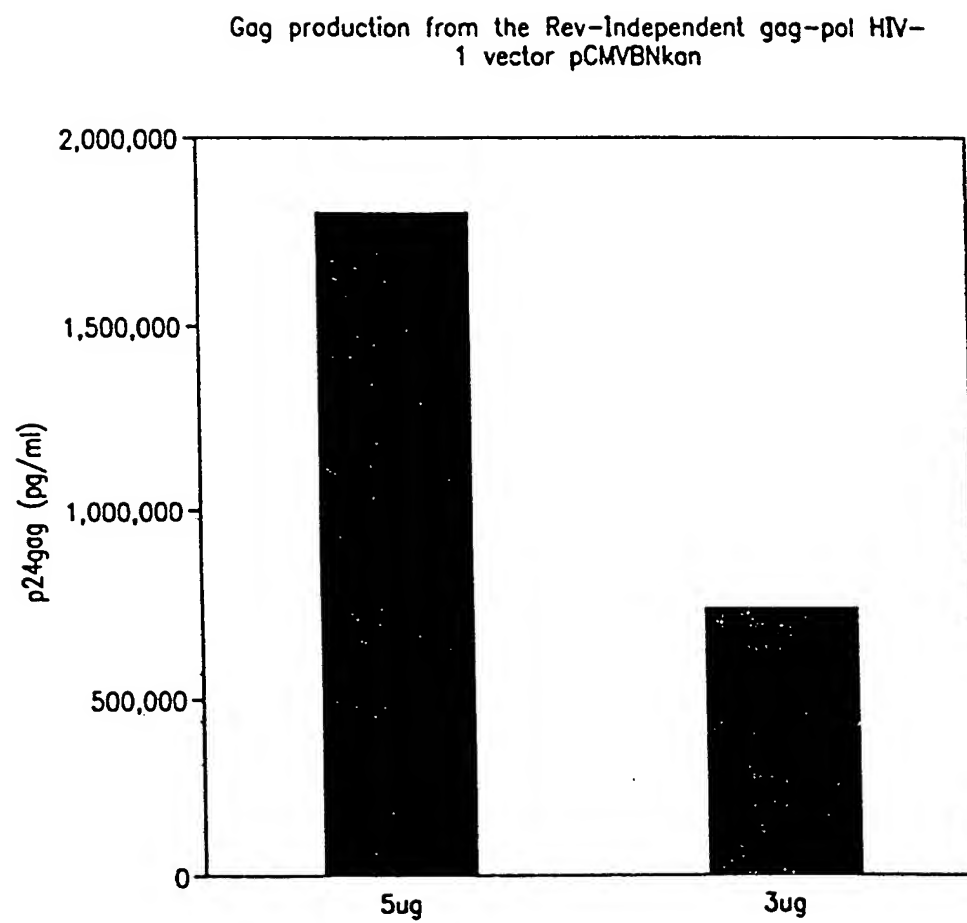


FIG. 13

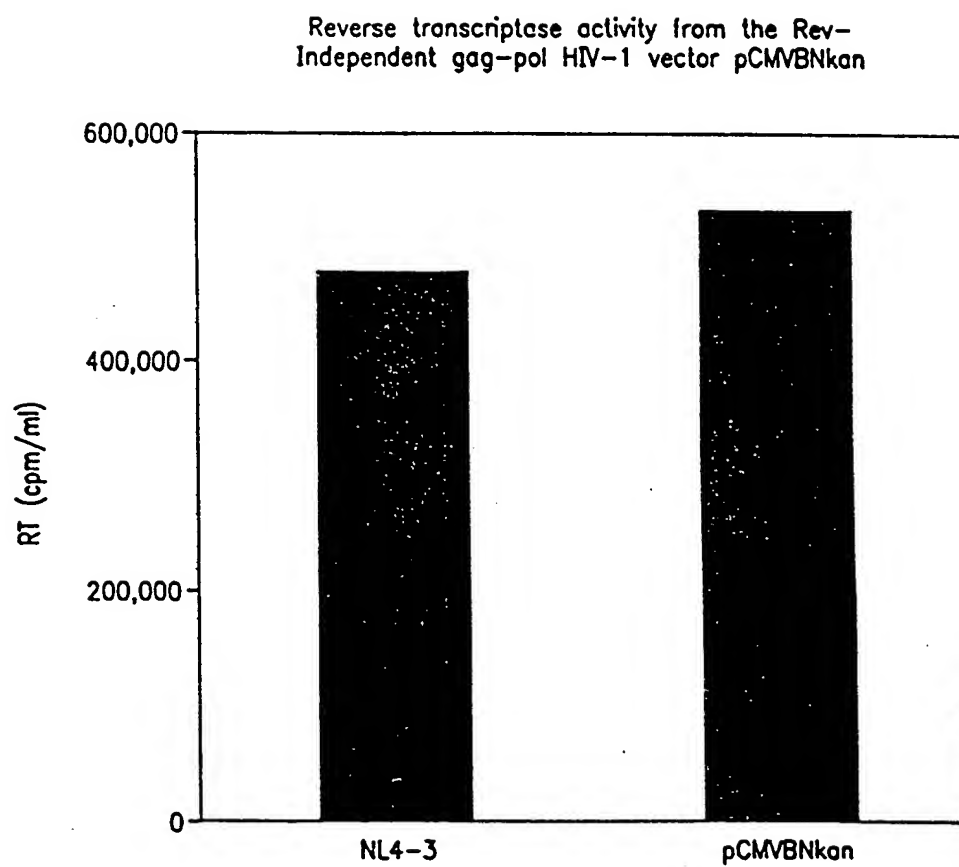


FIG. 14

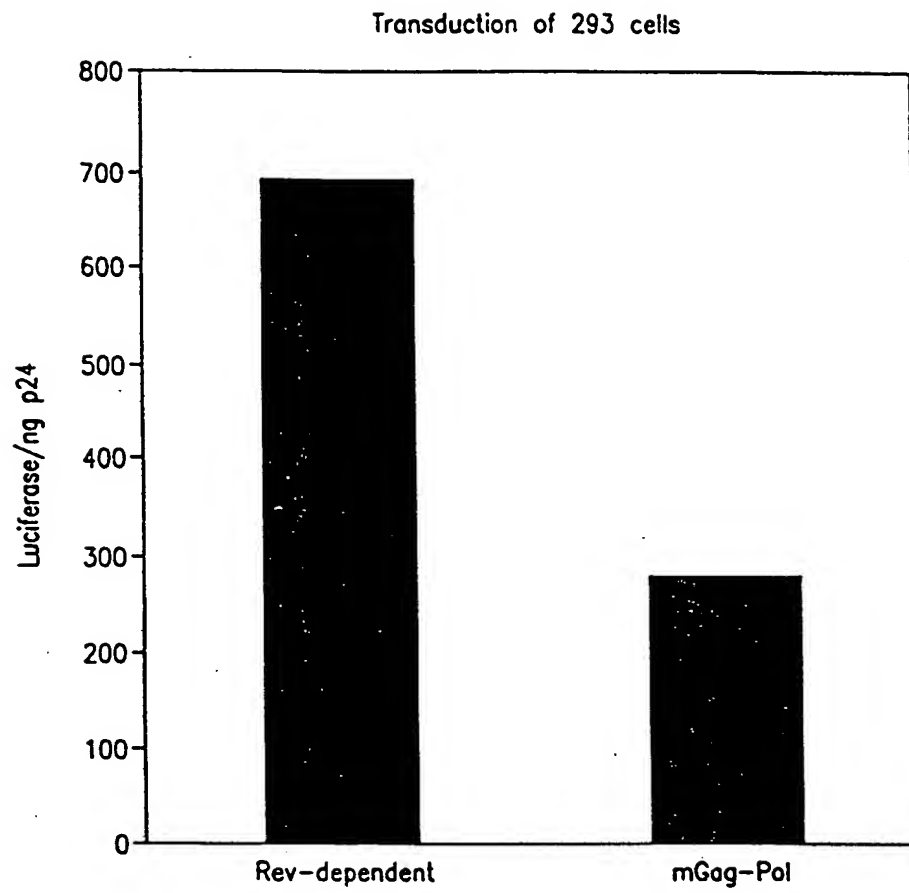


FIG. 15A

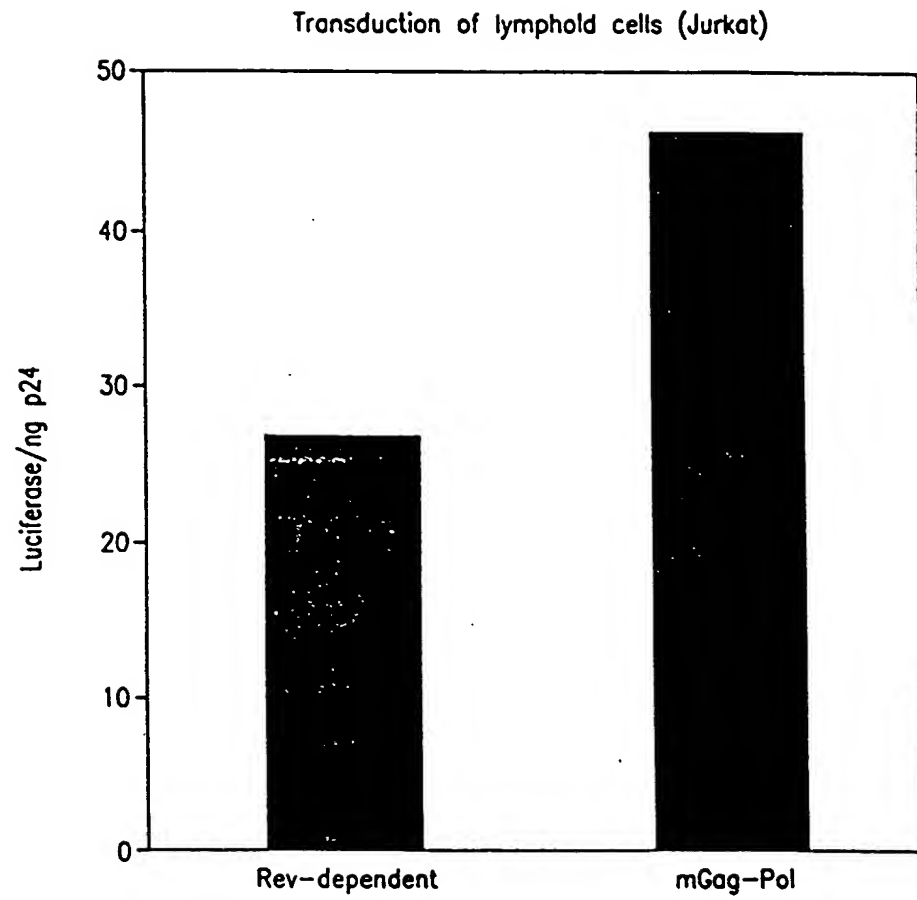


FIG. 15B

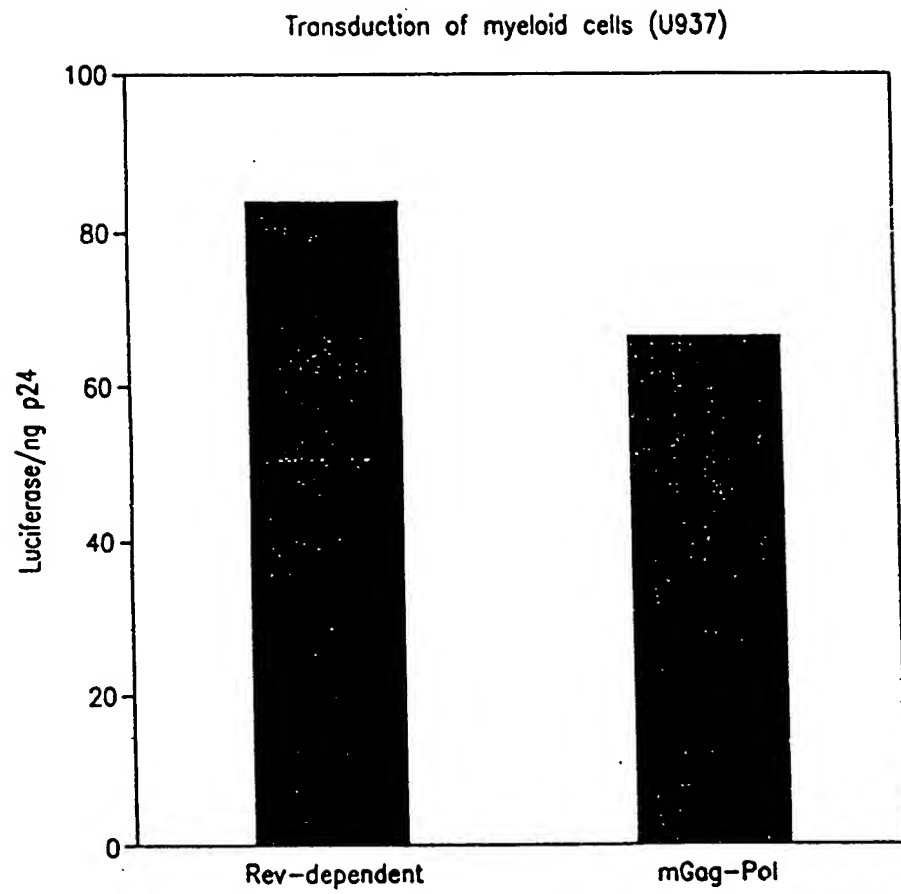


FIG. 15C

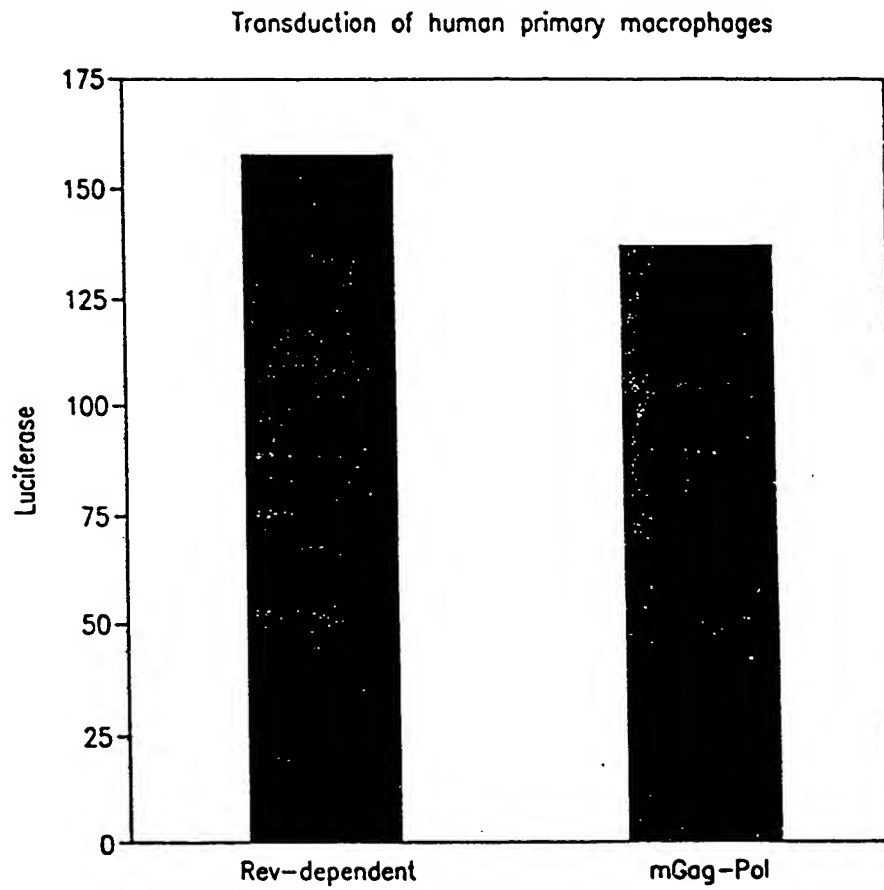


FIG. 15D

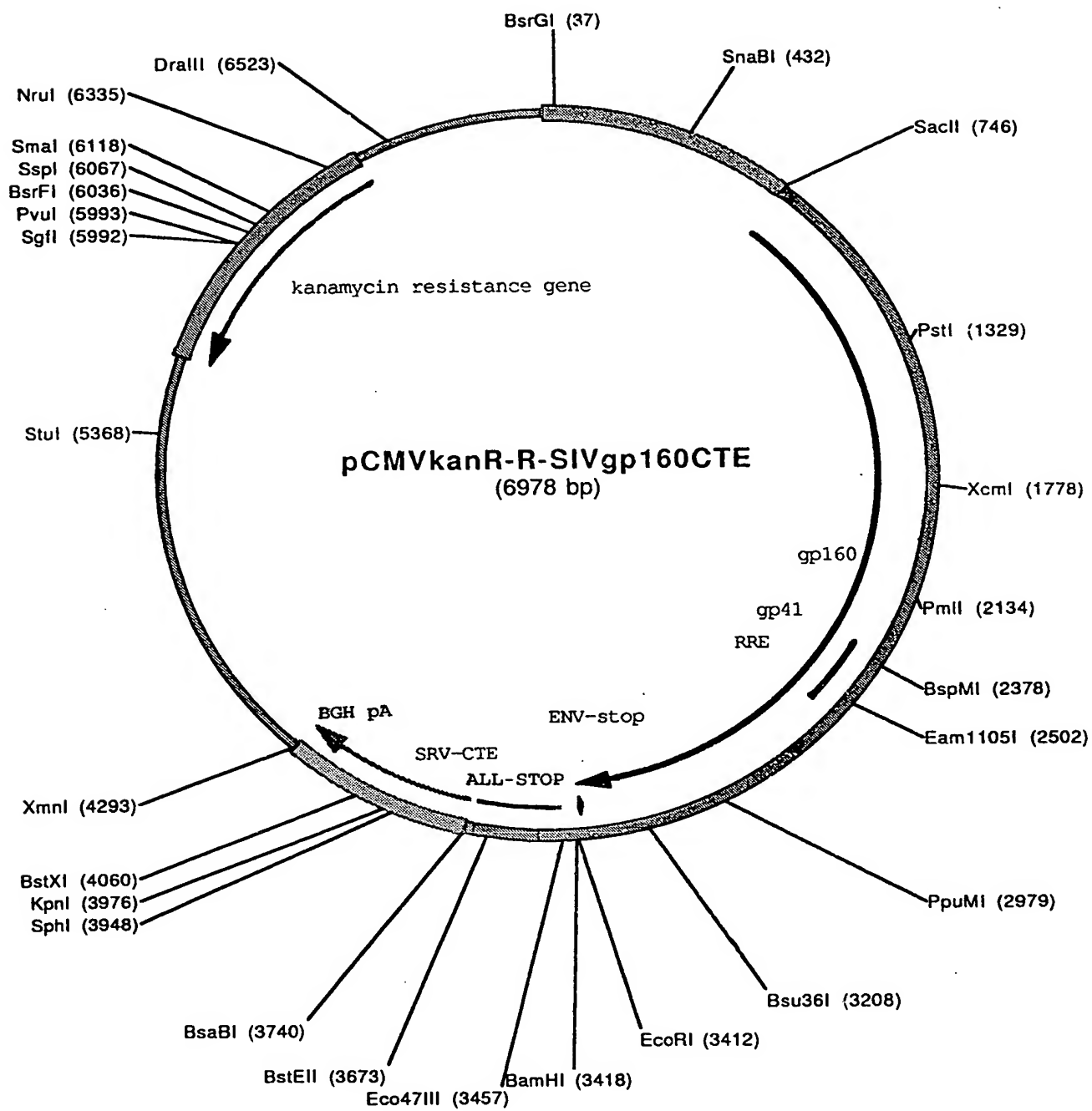


FIG. 16

42/45

BsrGI (37)  
 1 CCTGGCCATTGCATACGTTGTATCCATATCATAATGTACATTTATATGGCTCATGTCCAACATTACCGCCATGTTGA  
 81 CATTGATTATTGACTAGTTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTAC  
 161 ATAACCTACGGTAAATGGCCCGCCTGGCTGACCCCCCAACGACCCCGCCCATTTGACGTCAATAATGACGTATGTTCCCA  
 241 TAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA  
 321 GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGAC

SnaBI (432)  
 401 CTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTGGCAGTACA  
 481 TCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCOAAGTCTCAOCCCATTTGACGTCAATGGGAGTTTGTTTTGG  
 561 CACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTG  
 641 GGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCTGGAGACGCCATCCACGCTGTTTTGACCTCCATA

SacII (746)  
 721 GAAGACACCGGGACOGATCCAGCCTCCGCGGGCCGCGCTAAGTATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCT

1►Met Gl yCysLeu Gl yAsn Gl nLeuLeu l l eAl a l l eL  
 801 TGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACA

13►euLeuLeuSer Val Tyr Gl y l l eTyrCysThr LeuTyrVal Thr Val PheTyr Gl yVal ProAl aT rpArgAsnAl aThr  
 881 ATTCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACCTCAGTGCCTACCAGATAATGGTGATTATTGAGA

40►l l eProLeuPheCysAl aThr LysAsnArgAspThr Trp Gl yThr Thr Gl nCysLeuProAspAsn Gl yAspTyrSer Gl  
 961 AGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAGAGGATGTATGGCAAC

66►uValAl aLeuAsnVal Thr Gl uSer PheAspAl aT rpAsnAsnThr Val Thr Gl uGl nAl a l l eGl uAspVal Trp Gl nL  
 1041 TCTTTGAGACCTCAATAAAGCCTTGTTGTAATAATTATCCCCATTATGCATTACTATGAGATGCAATAAAGTGAGACAGAT

93►euPheGl uThr Ser l l eLysProCysVal LysLeuSer ProLeuCys l l eThrMetArgCysAsnLysSer Gl uThrAsp  
 1121 AGATGGGATTGACAAAATCAATAACAACACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAA

120►A rgTrp Gl yLeuThr LysSer l l eThr Thr Thr Al aSer Thr Thr Ser Thr Thr Al aSer Al aLysVal lAspMet Val lAs  
 1201 TGAGACTAGTTCTTGTATAGCCAGGATAATTGCACAGGCTTGGACAAGAGCAAAATGATAAGCTGTAAATCAACATGA

146►nGl uThr Ser Ser Cys l l eAl aGl nAspAsnCysThr Gl yLeuGl uGl nGl uGl nMet l l eSer CysLysPheAsnMet T  
 PstII (1329)  
 1281 CAGGGTTAAAAAGAGACAAGAAAAAGAGTACAATGAACTTGGTACTCTGCAGATTGGTATGTGAACAAGGGAATAAC

173►hr Gl yLeuLysArgAspLysLysLys Gl uTyrAsn Gl uThr TrpTyrSer Al aAspLeuVal Cys Gl uGl nGl yAsnAsn  
 1361 ACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGA

200►Thr Gl yAsn Gl uSer A rgCysTyrMetAsnHl sCysAsnThr Ser Val l l eGl nGl uSer CysAspLysHl sTyrTrpAs  
 1441 TGCTATTAGATTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATTATTAGGCTTTATGC

226►pAl a l l eArgPheArgTyrCysAl aProPro Gl yTyrAl aLeuLeuArgCysAsnAspThrAsnTyrSer Gl yPheMet P  
 1521 CTAAATGTTCTAAGGTGGTGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGA

253►roLysCysSer LysVal Val Val Ser Ser CysThr A rgMetMet Gl uThr Gl nThr Ser Thr TrpPhe Gl yPheAsn Gl y  
 1601 ACTAGAGCAGAAAATAGAATTATATTACTGGCATGGTAGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAA

280►Thr A rgAl aGl uAsnArgThr Tyr l l eTyrTrpHis Gl yA rgAspAsnArgThr l l e l l eSer LeuAsnLysTyrTyrAs  
 1681 TCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCAC

306►nLeuThrMetLysCysArgArgPro Gl yAsnLysThr Val LeuProVal Thr l l eMetSer Gl yLeuVal PheHisSer G  
 XcmI (1778)  
 1761 AACCAATCAATGATAGCCAAAGCAGGCATGGTGTGGTTTGGAGGAAATGGAAGGATGCAATAAAGAGGTGAAGCAG

333►l nPro l l eAsnAspArgProLys Gl nAl aT rpCysTrpPhe Gl yGl yLysTrpLysAspAl a l l eLys Gl uVal Lys Gl n  
 1841 ACCATTGTCAAACATCCAGGTATACTGGAATAACAATACTGATAAAATCAATTTGACGGCTCTGGAGGAGGAGATCC

360►Thr l l eVal LysHl sProArgTyrThr Gl yThrAsnAsnThrAspLys l l eAsnLeuThr Al aPro Gl yGl yGl yAspPr  
 1921 GGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTTCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAG

386►oGl uVal Thr PheMet TrpThrAsnCysArg Gl yGl uPheLeuTyrCysLysMetAsnTrpPheLeuAsnTrpVal Gl uA  
 2001 ATAGGAATACAGCTAACCAAGCCAAAGGAACAGCATAAAAGGAATTAOGTGCCATGTCATATTAGACAAATAATCAAC

413►spArgAsnThr Al aAsn Gl nLysProLys Gl uGl nHl sLysArgAsnTyrVal ProCysHis l l e Arg Gl n l l e l l eAsn

FIG. 17

2081 ACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCTCCAAGAGAGGGAGACCTCACGTGTAACCTCCACAGTGACCAGTCT  
PmlI (2134)

440 Thr TrpHisLysVal Gl yLysAsnVal TyrLeuProProArgGl uGl yAspLeuThr CysAsnSer Thr Val Thr Ser Le  
2161 CATAGCAAACATAGATTGGATTGATGGAACCAACTAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGG

466 u l l e A l a A s n l l e A s p T r p l l e A s p G l y A s n G l n T h r A s n l l e T h r M e t S e r A l a G l u V a l A l a G l u L e u T y r A r g L e u G  
2241 AATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCACAGATGTGAAGAGGTACACTACTGGTGGC

493 l u L e u G l y A s p T y r L y s L e u V a l G l u l l e T h r P r o l l e G l y L e u A l a P r o T h r A s p V a l L y s A r g T y r T h r T h r G l y G l y  
BspMI (2378)

2321 ACCTCAAGAAATAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCCTGCAATGGAGCGCG  
BspMI (2378)

520 Thr Ser ArgAsnLysArgGl yVal PheVal LeuGl yPheLeuGl yPheLeuAl aThr Al aGl ySer Al aMet Gl yAl aAl  
2401 CAGCCTGACCCCTACGGCACAGTCCCGAACTTTATTGGCTGGGATAGTCCAACAGCAGCAACAGCTGTTGGACGTGGTCA

546 aSer LeuThr LeuThr Al aGl nSer ArgThr LeuLeuAl aGl y l l eVal Gl nGl nGl nGl nGl nLeuLeuAspVal Val L  
Eam1105I (2502)

2481 AGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAGAACCCTCCAGACTAGGGTCACTGCCATCGAGAAGTAC

573 ysArgGl nGl nGl uLeuLeuArgLeuThr Val T rpGl yThr LysAsnLeuGl nThr ArgVal Thr Al a l l eGl uLysTyr  
2561 TTAAAGGACCAAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGC

600 LeuLysAspGl nAl aGl nLeuAsnAl aT rpGl yCysAl aPheArgGl nVal CysHis Thr Thr Val ProT rpProAsnAl  
2641 AAGTCTAACACCAAGTGGACAATGAGACTTGGCAAGAGTGGGAGCGAAAGTTGACTTCTTGAAGAAAATATAACG

626 aSer LeuThr ProLysT rpAsnAsnGl uThr T rpGl nGl uT rpGl uArgLysVal AspPheLeuGl uGl uAsn l l eThr A  
2721 CCCTCTAGAGGAGGCACAAATTCACAAGAGAAGAACATGTATGAATTACAAAGTTGAATAGCTGGGATGTGTTTGGC

653 l aLeuLeuGl uGl uAl aGl n l l eGl nGl nGl nLysAsnMet TyrGl uLeuGl nLysLeuAsnSer T rpAspVal PheGl y  
2801 AATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAAT

680 AsnT rpPheAspLeuAl aSer T rp l l eLysTyr l l eGl nTyrGl yVal Tyr l l eVal Val Gl yVal l l eLeuLeuArg l l  
2881 AGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGTATAGGCCAGTGTCTCTTCCCAACCTCTTATTTC

706 eVal l l eTyr l l eVal Gl nMetLeuAl aLysLeuArgGl nGl yTyrArgProVal PheSer Ser ProProSer TyrPheG  
PpuMI (2979)

2961 AGCAGACCATATCCAACAGGACCCGGCACTGCCAACAGAGAAGGCAAGAAAGAGACGGTGGAGAAGCGGTGGCAAC

733 l nGl nThr His l l eGl nGl nAspProAl aLeuProThr ArgGl uGl yLysGl uArgAspGl yGl yGl uGl yGl yAsn  
3041 AGCTCTGGCCTTGGCAGATAGAATATATCCACTTCTTATTGTCAGCTTATTAGACTCTTGACTTGGCTATTTCAGTAA

760 Ser Ser T rpProT rpGl n l l eGl uTyr l l eHisPheLeu l l eArgGl nLeu l l eArgLeuLeuThr T rpLeuPheSerAs  
3121 CTGTAGGACTTTGCTATCGAGAGTATACAGATCTCCAACCAATACTCCAGAGGCTCTCTCGACCCCTACAGAGGATTC

786 nCysArgThr LeuLeuSer ArgVal TyrGl n l l eLeuGl nPro l l eLeuGl nArgLeuSer Al aThr LeuGl nArg l l eA  
Bsu36I (3208)

3201 GAGAAGTCTCAGGACTGAAGTACCTACCTACAATATGGGTGGAGCTATTTCATGAGGGGTCCAGGCCGTCTGGAGA

813 r gGl uVal LeuArgThr Gl uLeuThr TyrLeuGl nTyrGl yT rpSer TyrPheHisGl uAl aVal Gl nAl aVal T rpArg  
3281 TCTGCGACAGAGACTCTTGGGGCGCGTGGGAGACTTATGGGAGACTCTTAGGAGAGGTGGAAGATGGATACTCGCAAT

840 Ser Al aThr Gl uThr LeuAl aGl yAl aT rpGl yAspLeuT rpGl uThr LeuArgArgGl yGl yA r gT r p l l eLeuAl a l l  
BamHI (3418)

3361 CCCCAGGAGGATTAGACAAGGGCTTGAGCTCACTCTCTTGTGAGGGACAGAGAATTCGGATCCactagtctctagaCTOGA  
EcoRI (3412)

866 eProArgArg l l eArgGl nGl yLeuGl uLeuThr LeuLeu . . .  
Eco47III (3457)

3441 GGGGGGGCCCGGTACGAGCGCTTAGCTAGCTAGAGACCCTCCCTGCGAGCTAAGCTGGACAGCCAATGACGGGTAAG

3521 AGAGTGACATTTTCTACTAACCTAAGACAGGAGGGCGTCAGAGCTACTGCCTAATCCAAGACGGGTAAAAGTGATAAA

3601 AATGTATCACTCCAACCTAAGACAGGCGCAGCTTCCGAGGGATTTGTCGCTGTTTTATATATATTAAAGGGTGACCT  
BstEII (3673)

3681 GTCCGGAGCCGTGCTGCCCGGATGATGTCTTGGTCTAGACTCGAGGGGGGGCCCGGTACGATOCAGATCTGTGTGCTT  
 3761 CTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCTTCTTGACCTGGAAGGTGCCACTCCCACTGTCCTTTCC  
 3841 TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCTATTCTATTCTGGGGGGTGGGGTGGGACAGCACAGCAA  
 SphI (3948) KpnI (3976)  
 3921 GGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGATGCGGTGGGCTCTATGGGTACCGAGGTGCTGAAGAATTGAC  
 BstXI (4060)  
 4001 CCGGTTCTCTCTGGGCCAGAAAGAAGCAGGCACATCCCTTCTCTGTGACACACCTGTCCACGCCCTGGTTCTTAGTT  
 4081 CCAGCCCCACTCATAGGACACTCATAGCTCAGGAGGGCTCCGCTTCAATCCACCGCTAAAGTACTTGGAGCGGTCTC  
 4161 TCCCTCCCTCATCAGCCACCAACCAACCTAGCCTCCAAGAGTGGGAAGAAATTAAAGCAAGATAGGCTATTAAGTGC  
 XmnI (4293)  
 4241 AGAGGGAGAGAAAATGCCTCCAACATGTGAGGAAGTAATGAGAGAAATCATAGAATTTCTTCCGCTTCTCGCTCACTGA  
 4321 CTCGCTGCGCTCGGTGCTTCCGCTGCGGCGAGCGGTATCAGCTCACTCAAAGCGGTAATACGGTTATCCACAGAATCAG  
 4401 GGGATAACGCAGGAAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAGGCCCGCGTTGCTGGCGTTT  
 4481 TTCCATAGGCTCCGCCCCCTGAAGAGCATCAAAAAATCGACGCTCAAGTCAGAGGTGGGAAACCCGACAGGACTATA  
 4561 AAGATACCAGCGCTTCCCGCTGGAAGCTCCCTCGTGGCTCTCTGTTCCGACCCCTGCCGCTTACCGGATACCTGTCCG  
 4641 CCTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCAAGCTGTAGGTATCTCAGTTCCGTTGAGGTGCTTCGCTCC  
 4721 AAGCTGGGCTGTGTGCACGAACCCCGCTTCAGCCCGACCGCTGCGCTTATCCGTTAACTATCGTCTTGAAGTCAACCC  
 4801 GGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAG  
 4881 AGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTGGTATCTGCGCTCTGCTGAAGCCAGTTACC  
 4961 TTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAACAAACCCCGCTGGTAGCGGTGGTTTTTTTGTGTCAGCAGCA  
 5041 GATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAACT  
 5121 CACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTAAAAATGAAGTTTAAA  
 5201 TCAATCTAAAGTATATATAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTG  
 5281 TCTATTTCGTTTCATCCATAGTTGCCTGACTCCGGGGGGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGAC  
 StuI (5368)  
 5361 TCATACCAGGCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCAGGTTGATGAGAGCTTTGTTGTAGGTGGA  
 5441 CCAGTTGGTGATTTTGAACTTTGTCTTGCCACGGAACGGTCTGCGTTGTGCGGAAGATGCGTGATCTGATCCTTCAACT  
 5521 CAGCAAAAGTTGATTTATTCAACAAAGCCCGCTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAA  
 5601 CCAATCTGATTAGAAAACTCATCGAGCATCAAAATGAAGTCAATTTATTCATATCAGGATTATCAATACCATATTTT  
 2714 PhePheGluAspLeuMetLeuHisPheGlnLeuLysAsnMetAspProAsnAspIleGlyTyrLysG  
 5681 TGAAAAAGCCGTTTCTGTAATGAAGGAGAAACTCACGAGGCAGTTCATAGGATGGCAAGATCCTGGTATCGGTCTGC  
 2484 InPheLeuArgLysGlnLeuSerProSerPheGluGlyLeuCysAsnTrpLeuIleAlaLeuAspGlnTyrArgAspAla  
 5761 GATTCGAGCTCGTCCAAATCAATACAACTATTAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCAACAT  
 2224 IleGlyValAlaArgGlyValAspIleCysGlyIleLeuLysGlyGluAspPheIleLeuAsnAspLeuSerPheAspGlyHis  
 5841 GAGTGACGACTGAATCCGGTGAGAAATGGCAAGCTTATGCATTTCTTCCAGACTGTTCACAGGCCAGCCATTACGC  
 1954 sThrValValSerAspProSerPheProLeuLeuLysHisMetGluLysTrpValGlnGluValProTrpGlyAsnArgG  
 PvuI (5993)  
 5921 TCGTCATCAAAATCACTCGCATCAACCAACCGTTATTCATTCGTGATTGCGCTGAGCGAGACGAAATACGGATCGCT  
 1684 LuAspAspPheAspSerAlaAspValLeuGluAsnAsnMetArgSerGlnAlaGlnAlaLeuArgPheValAlaArgAspSer  
 BsrFI (6036) SspI (6067)  
 6001 GTTAAAGGACAATTACAAACAGGAATCGAATGCAACCGCGCAGGAACACTGCCAGCGCATCAACAAATATTTTACCTG  
 1424 AsnPheProCysAsnCysValProIleSerHisLeuArgArgLeuPheValAlaLeuAlaAspValIleAsnGluGlySe  
 SmaI (6118)  
 6081 AATCAGGATATCTTCTAATACTGGAATGCTGTTTTCCCGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTA  
 1154 rAspProTyrGluGluLeuValGlnPheAlaThrLysGlyProIleAlaThrThrLeuLeuTrpAlaAspAspProThrA  
 6161 CGGATAAAATGCTTGATGGTGGGAAGAGGCATAAATCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATT  
 884 rGllePheHisLysIleThrProLeuProMetPheGluThrLeuTrpAsnLeuArgValMetGluAspThrValAspAsn  
 6241 GGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCGGCGCATCGGGCTTCCCATACAATCGATAGATTGTGCGACCTG  
 624AlaValSerGlyLysGlyHisLysLeuPheLeuGluProAlaAspProLysGlyTyrLeuArgTyrIleThrAlaGlySe  
 NruI (6335)  
 6321 ATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCTCGAGCAA  
 354 rGlnGlyValAsnAspArgAlaTrpLysTyrGlyTyrLeuAspAlaAspMetAsnSerAsnLeuArgProArgSerCysS  
 6401 GACGTTCCCGTTGAATATGGCTCATACACCCCTTGATTACTGTTATGTAAGCAGACAGTTTATTGTTTCATGATGA  
 84 rThrGluArgGlnIleHisSerMet  
 DraIII (6523)  
 6481 TATATTTTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCCCCCCCCCCCATTATTGAAGCA  
 6561 TTTATCAGGGTTATTGTCTCATGAGCGGATACATATTGAATGTAATTAGAAAAATAAACAATAGGGGTTCCGCGCACA

45/45

6641 TTTCCCGAAAAGTGCCACCTGACGTCTAAGAAAACATTATTATCATGACATTAACCTATAAAAAATAGGCGTATCACGAG  
6721 GCCCTTTTCGTC TC GCGCGTTTGGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCGGAGACGGTCACAGCTTGTC  
6801 TGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACAT  
6881 GCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAAATACCG  
6961 CATCAGATTGGCTATTGG